BLENDED CMA: SMOOTH, ADAPTIVE TRANSFER FROM CMA TO DD-LMS

L. R. Litwin, Jr.*, M. D. Zoltowski*†, T. J. Endres‡, and S. N. Hulyalkar‡

Purdue University* and Sarnoff Digital Communications‡

ABSTRACT

A common technique in a blind equalization communication system is to use CMA to open the channel eye and then to switch to decision-directed LMS once the error rate has dropped to a sufficiently low level. We propose a new algorithm, Blended CMA, which dynamically selects to use either the CMA or DD-LMS update term when adapting the equalizer taps. The decision rule governing this selection is also adaptive. Simulation results presented in this paper demonstrate how the algorithm performs when equalizing 64-QAM over real-world channels and highlights the performance gain over using traditional transfer methods.

1. INTRODUCTION

In this paper we propose a new algorithm named Blended CMA (B-CMA) which uses adaptable-radius decision circles to determine whether to update the equalizer using CMA [1, 2] or DD-LMS [3]. The motivation for this algorithm was to obtain a lower error rate at the time of transfer to DD-LMS. Examples of other work in this area can be found in [4] and [5]. The B-CMA algorithm that we propose does not make an abrupt transfer to DD-LMS, but instead it gradually makes the transfer. It does this by deciding to use DD-LMS only if the current equalizer output falls within one of the decision circles that are centered at each constellation point. Based on an easily computed metric, the radius of the decision circles is increased gradually as the equalizer output begins to converge around each constellation point. In order for the algorithm to be robust to "poor" choices for the parameters that determine when the radius expands, B-CMA contains logic that adaptively adjusts these parameters and hence the algorithm "teaches" itself good parameters to use. Simulation results will be presented that compare the performance of B-CMA to that of CMA using 64-QAM over real-world channels. The simulations demonstrate how B-CMA obtains a significantly lower error rate at the time of transfer to DD-LMS compared to using standard CMA. A more complete discussion of B-CMA can be found in [6].

The rest of this paper is organized as follows. Section 2 briefly discusses CMA and DD-LMS. Section 3 presents the proposed Blended CMA algorithm. Section 4 describes the assumptions and settings used for the simulations. Section 5 presents simulation results. Section 6 provides concluding remarks.

2. CMA AND DD-LMS

CMA is commonly used as a cold start-up algorithm in situations where severe ISI causes the channel eye to be closed and it is not initially possible to use DD-LMS due to the high number of incorrect decisions. CMA is a stochastic gradient descent algorithm and its tap update equation is written as

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \mathbf{r}_n^* \underbrace{y_n(\gamma - |y_n|^2)}_{CMA\ error} \tag{1}$$

where μ is a tunable, positive constant called the step-size and γ is a positive constant known as the Godard radius. y_n represents the equalizer output at baud instance n, $\mathbf{f}_n$ is the length-M vector of equalizer taps, and $\mathbf{r}_n$ is the regressor vector formed from the M most recent equalizer input samples. Complex conjugation is denoted by the asterisk.

DD-LMS is also a stochastic gradient descent algorithm and its tap update equation is written as

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \mathbf{r}_n^* \underbrace{(\hat{s}_n - y_n)}_{LMS\ error} \tag{2}$$

where $\hat{s}_n$ is the hard decision that is formed by the receiver's decision device. CMA can be used to update the taps until the equalizer has converged to the point that the transfer to DD-LMS can be made reliably. Practitioners define the condition for reliable transfer to DD-LMS in terms of a symbol error rate (SER) between 10^{-1} and 10^{-2} [7]. We will approximate this condition as an SER of 0.04. For 64-QAM

* Department of Electrical Engineering, Purdue University, West Lafayette, IN 47906, litwin(mikedz)@ecn.purdue.edu

† Supported in part by the National Science Foundation under grant no. MIPS-9708309.

‡ Sarnoff Digital Communications, Suite 100, 6 Penns Trail, Newtown, PA 18940, endres(samirh)@sdcomm.com

WCNC, New Orleans 1999

and a normalized channel as defined in [7], this SER corresponds to a mean-squared error (MSE) of approximately 0.0182 (see [8] for the underlying mathematics). We will use the term Transfer Level to refer to this MSE value. Thus, CMA is used until the MSE falls below the Transfer Level at which point the transfer to DD-LMS is made.

3. BLENDED CMA

3.1. Motivation

The development of B-CMA was motivated by the desire to increase the reliability of the transfer from CMA to DD-LMS. A summary in [9] discusses CMA's robustness properties to practical situations by describing the proximity of the CM local minima to the Wiener settings. However, the transfer from CMA to DD-LMS is often unreliable due to the increase in the stochastic jitter, or Excess MSE (EMSE), that occurs when using CMA with a non-CM source (i.e., higher order QAM).

A plot (see [6]) of the cubic CMA error term in (1) for a non-CM source reveals that the CMA error term is non-zero at the symbol points. Thus, the taps will continue to adapt even when the equalizer output lies on a symbol point. The misadjustment factor due to this adaptation will increase the EMSE. In contrast to CMA, the DD-LMS error term in (2) assumes zero values at the symbol points and hence the EMSE is greatly reduced. B-CMA takes advantage of DD-LMS's reduced EMSE by adaptively selecting between CMA and DD-LMS. The choice is derived from thresholding the DD-LMS error term. Small values of $|\hat{s}_n - y_n|$ correspond to circular regions centered at a symbol point. For these situations the taps are updated with DD-LMS. Larger values of $|\hat{s}_n - y_n|$ correspond to regions that are far from a symbol point and thus CMA is used.

3.2. The Blended CMA Algorithm

The B-CMA algorithm makes use of adaptable-radius decision circles to determine whether to use the CMA or DD-LMS tap update term to adapt the equalizer taps. The algorithm determines which update term to use based on where the current equalizer output lies. Identical decision circles, with adaptable radius R, are centered around each constellation point and if the current equalizer output falls within one of these decision circles, the equalizer updates the taps using the DD-LMS tap update term. If the symbol falls outside of these decision circles, then the taps are updated with the CMA tap update term. The reasoning for this approach is that symbols that are close to a given constellation point are assumed to correspond to that constellation point, and hence the resulting hard decision is assumed to be correct and DD-LMS is used. However, for regions further from constellation points, it is more likely that the corresponding hard decision will be incorrect and thus it should not be used for a decision-directed calculation. Since CMA relies on the statistics of the signal as opposed to hard decisions, it can be used to compute the tap update term when symbols lie farther away from constellation points.

The key to B-CMA's performance is its ability to adapt the radius of the decision circles. This adaptation allows B-CMA to use decision-directed updates more frequently as the channel eye is opened. In addition, if a time-varying channel causes the eye to close, B-CMA uses CMA as an aquisition aid in order to reopen the eye. Figure 1 shows the decision regions for a single constellation point, and the figure is useful for understanding how the adaptation process works. The dashed circle of radius R is the decision circle and the value for R is adapted by the algorithm as symbols are processed. A monotonically increasing sequence of values used for R is stored in the vector $\mathbf{R}$ and the values lie within the range $0 \leq \mathbf{R}(i) \leq 1$. The index i denotes the current location in the vector. The dotted circle of radius $R_{th} = 1$ is the threshold circle. The radius R is initially zero and thus the equalizer is updated using CMA. The symbol counter k is incremented as each symbol is processed. No radius adaptation occurs until at least N symbols have been processed, i.e. until $k \geq N$. At that point, the algorithm computes the number of the previous N symbols that were within a distance of $R_{th} = 1$ from their corresponding hard decisions. This number is assigned to *sum*. The value of *sum* is then compared to a threshold. A set of thresholds is stored in the vector $\mathbf{T}$, and each element in the threshold vector is related to an element in $\mathbf{R}$. Hence, if the current decision radius is $\mathbf{R}(i)$, the current value of the threshold is $\mathbf{T}(i)$. A key feature of B-CMA is that the values of $\mathbf{T}$ are adapted by the algorithm. The adaptation is important because it allows the algorithm to adapt to changing channel characteristics and it also enables B-CMA to be robust to "poor" initial values of $\mathbf{T}$. The computed value of *sum* is compared to the threshold $\mathbf{T}(i)$. The decision radius $R = \mathbf{R}(i)$ is increased if $sum \geq \mathbf{T}(i)$. This increase is accomplished by incrementing i by one since the values stored in $\mathbf{R}$ are monotonically increasing. If R is already one, then B-CMA instead transfers to DD-LMS.

The values for $\mathbf{T}$ are user-defined, and since the optimal (in some sense) values vary among channel conditions, we desire an adaptive method of choosing the parameter threshold vector $\mathbf{T}$ based on a suitable initialization. For example, if an element in $\mathbf{T}$ is set too

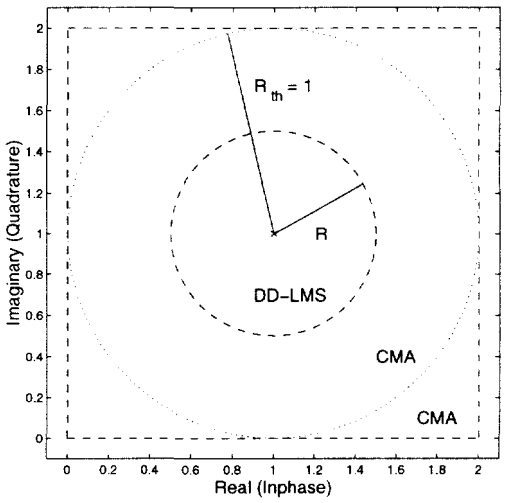


Figure 1: Example of a B-CMA decision region. The square region represents the standard DD-LMS decision region and the circles represent the decision (R) and threshold (R_{th}) circles used by B-CMA.

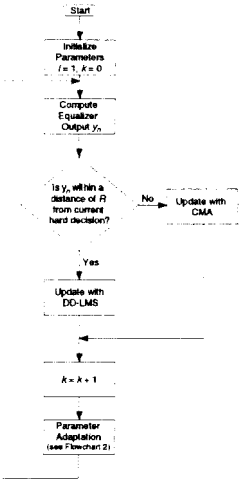


Figure 2: Flowchart of main loop of B-CMA algorithm.

low, the decision radius will be increased prematurely and the algorithm performance may suffer due to the increased number of incorrect decisions. Fortunately, B-CMA adapts the threshold vector in response to this increase in incorrect decisions. If *sum* was not greater than the current value of the threshold, $\mathbf{T}(i)$, the algorithm then checks to see if *sum* is less than the previous value of the threshold, $\mathbf{T}(i-1)$. If it is, the algorithm determines that R is set too high and it decreases the decision radius by decrementing i by one. After i is decremented, the algorithm then increases the threshold $\mathbf{T}(i)$ by T_{inc}. A recommended value for T_{inc} is about 2-4% of N. Since the decision radius was assumed to be increased prematurely, the threshold is increased by T_{inc} in order to force the algorithm to wait longer before switching back to that same decision radius. Thus, if the values of $\mathbf{T}$ are set too low, B-CMA will adapt the values until it finds good settings itself.

Alternatively, it is possible that threshold values can be set too high. In such a case, the value of *sum* would not be large enough for adaptation of R to occur. If $5N$ symbols have been processed without any adaptation, the algorithm decreases the value of the current threshold $\mathbf{T}(i)$ by T_{leak}. A recommended value for T_{leak} is about 1-2% of N. By decreasing the threshold slightly, the algorithm is able to handle the situation where the threshold values have been set at a level that is too high for adaptation to occur.

The mechanics of the B-CMA algorithm are summarized in Figures 2 and 3. Figure 2 describes the logic of the main loop of the algorithm and the parameter adaptation is shown separately in Figure 3.

4. ASSUMPTIONS/SETTINGS

4.1. Assumptions

All data was fractionally sampled with a spacing of $T/2$ where T is the symbol period. The channel model is derived from experimentally acquired microwave data and is available at the SPIB database[1]. The simulation results from channel 10 are shown in this paper, however, more results are available in [6]. The results presented are for the case of white Gaussian noise with an SNR of 40 dB. The received signal variance was set to unity using the normalization described in [7].

4.2. Settings

The simulation used a 32-tap equalizer and the value of the LMS step-size was $\mu_{LMS} = 0.0200$. The two values for the CMA step-size were $\mu_1 = 0.0020$ and $\mu_2 = 0.0005$. The larger μ_1 was also used by the B-CMA algorithm when updating the taps with CMA. The taps were initialized with the center tap set to unity and all others zeroed.

The B-CMA algorithm settings that were used for the simulation were $N = 100$, $\mathbf{T} = [70\ 75\ 80\ 85\ 90\ 95\ 100]$ and $\mathbf{R} = [0.0\ 0.1\ 0.25\ 0.35\ 0.5\ 0.75\ 1.0]$. The parameter $\mathbf{T}(7) = 100 = N$ represents the threshold for the equalizer to transfer to DD-LMS. Although it was included as a recommended setting for actual use,

[1]The Rice University SPIB channel database resides at http://spib.rice.edu/spib/microwave.html.

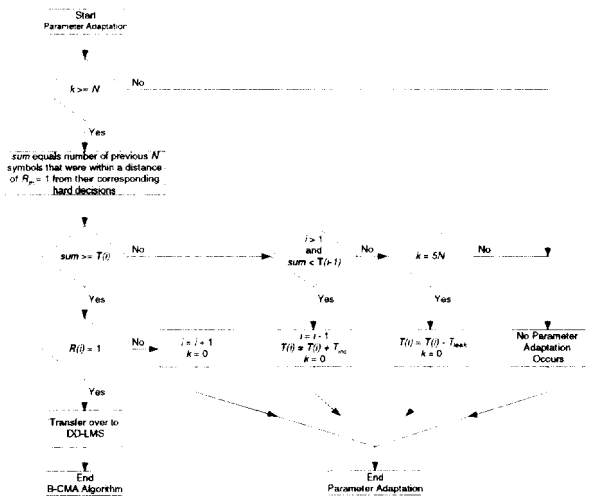


Figure 3: Flowchart of parameter adaptation loop of B-CMA algorithm.

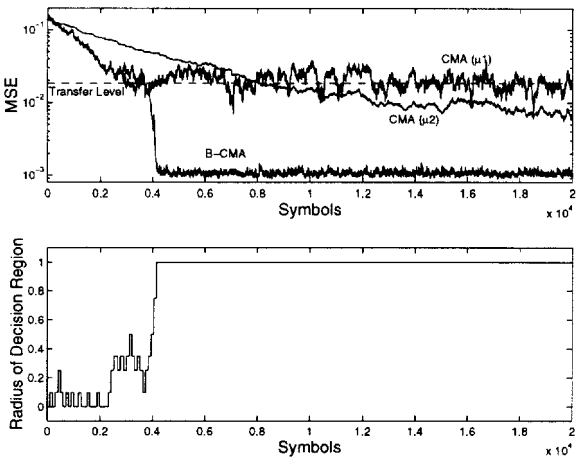


Figure 4: Simulation results for channel 10 with an SNR of 40 dB.

B-CMA was not allowed to transfer to DD-LMS for the simulation results shown in this paper. The transfer was prevented in order to compare the behavior of CMA and B-CMA prior to the transfer to DD-LMS.

5. SIMULATION RESULTS

The MSE trajectories for CMA and B-CMA are shown in Figure 4. CMA (μ_1) was able to get near the Transfer Level, however, the EMSE due to the large step-size prevented it from dropping below the Tranfer Level. The smaller step-size of CMA (μ_2) reduced the EMSE and enabled it to drop below the Transfer Level, however, the smaller step-size also slowed convergence. The step-size used by B-CMA (in CMA mode) was the same as CMA (μ_1). Although CMA performed poorly with μ_1, B-CMA was able to tolerate a higher μ_{CMA} due to the reduced EMSE of DD-LMS. B-CMA reached the Transfer Level quickly and its MSE dropped much lower compared to CMA. Thus, B-CMA was able to obtain a more reliable transfer point than CMA.

6. CONCLUSIONS

In this paper we presented a novel algorithm for cold start-up applications named Blended CMA. Simulation results were shown that demonstrated how B-CMA not only obtained a significantly lower MSE, but it also has the added benefit of converging faster than CMA. Thus, it reached a more reliable transfer point from which the equalizer could transfer to DD-LMS mode.

7. REFERENCES

[1] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," IEEE Transactions on Communications, vol. 28, no. 11, pp. 1867-1875, Nov. 1980.

[2] J. R. Treichler, B. G. Agee, "A new approach to multipath correction of constant modulus signals," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-31, no. 2, pp. 459-72, Apr. 1983.

[3] B. Widrow, M. Hoff, "Adaptive Switching Circuits," IRE WESCON Conv. Rec., pt. 4, pp. 96-104, 1960.

[4] V. Weerackody, S. A. Kassam, "Dual-mode type algorithms for blind equalization," IEEE Transactions on Communications, vol. 42, no. 1, pp. 22-28, Jan. 1994.

[5] Y. Lee, S. Kwon, S. Kang, K. Kim, C. Kang, "A decision-directed blind equalization with the error variance estimation," IEEE Intern. Conf. on Universal Personal Comm., vol. 1, pp. 99-103, Oct. 1997.

[6] L. R. Litwin, Jr., M. D. Zoltowski, T. J. Endres, S. N. Hulyalkar, "Blended CMA: Smooth, Adaptive Transfer from CMA to DD-LMS," submitted to IEEE Transactions on Communications, 1999.

[7] T. J. Endres, et al., "Simulated comparisions of blind equalization algorithms for cold start-up applications," Intern. Journal of Adaptive Control & Sig. Proc., vol. 12, no. 3, pp. 283-301, May 1998.

[8] J. G. Proakis, Digital Communications, New York, NY: McGraw-Hill, third edition, 1995.

[9] C. R. Johnson, Jr., et al., "Blind equalization using the Constant Modulus criterion: A review," Proceedings of the IEEE, vol. 86, no. 10, pp. 1927-1950, Oct. 1998.

A DECISION-DIRECTED CONSTANT MODULUS ALGORITHM FOR HIGHER-ORDER SOURCE CONSTELLATIONS

T. J. Endres, S. N. Hulyalkar, C. H. Strolle, T. A. Schaffer and R. A. Casas

NxtWave Communications
One Summit Square
Langhorne, PA 19047
{*endres,samirh,cstrolle,schaffer,raul*}*@nxtwavecomm.com*

ABSTRACT

This paper discusses methods for calculating and implementing the update error term for the popular blind equalization algorithm known as Godard's algorithm, or the Constant Modulus Algorithm (CMA), without the use of multipliers so that chip area and signal latency are both substantially reduced. A decision-directed CMA update term is derived for higher-order (non-constant modulus) source alphabets. The modified update error term can be calculated using a look-up table in place of costly multipliers and adders. Baseband and passband implementations for one-dimensional and two-dimensional signaling are discussed.

1. INTRODUCTION

Numerous consumer and government products relying on digital transmission in the new millenium require adaptive equalizers to restore corrupted data to acceptable error rates. The data rates used by current and next generation hardware are often so high that the equalizer operates in a data-rich environment so that methods relying on higher-order statistics, which were once thought too slow to be practical, can now provide robust demodulation with extremely fast acquisition time.

The Constant Modulus Algorithm (CMA) accumulates a third-order moment of the equalizer output. CMA was originally proposed by Godard [6] for QAM signals and developed independently by Treichler and Agee [12] for constant envelope FM signals. Some recent work focusing on reduced-complexity implementation of the CMA update propose a signed-error CMA (SE-CMA) which replaces the usual CMA error term with its signum (+/-1) [1], [5]. In [1], this 1-bit quantization is shown to significantly distort the Constant Modulus (CM) cost surface by making it piecewise flat, or tiled. Since the usual CM cost surface is multi-modal, the SE-CM cost surface can be composed of facets with zero slope so that equalizer converge can be indefinitely prolonged. Schniter and Johnson [11] attempt to remedy this potential problem by introducing a tunable dithering term to the CMA error term before signum operation. Hence, on average, over a certain region, the SE-CMA trajectory follows the true CMA trajectory. The dithering preserves CMA's robustness properties at the expense of raising the misadjustment, or excess MSE (EMSE).

Our work describes methods for quantizing the CMA error term so that it can be calculated using a look-up table in place of costly multipliers and adders, thereby reducing chip area and signal latency. Our methods are designed to provide better performance than the 1-bit quantization studied in [1] and [5], and can conceivably be combined with the dithering proposed in [11]. Both one-dimensional Pulse Amplitude Modulation (PAM) and two-dimensional Quadrature Amplitude Modulation (QAM) signaling are used as examples.

Following this introduction, §2 briefly reviews the Constant Modulus (CM) criterion and CMA. §3 describes the calculation of a decision-directed CMA (DD-CMA) error term for real-valued signal processing. §4 describes a DD-CMA error term for complex-valued signaling. §5 describes quantization using the DD-CMA error term when the equalizer processes passband samples which are not strictly at DC. §6 illustrates a numerical simulation, and §7 contains concluding remarks. The material discussed in this paper is patent pending [2].

2. CONSTANT MODULUS ALGORITHM PRIMER

The Constant Modulus (CM) cost function [6] is given by

$$J_{CM} = E\{(\gamma^2 - |y(k)|^2)^2\} \tag{1}$$

where γ^2 is a scalar referred to as Godard's (dispersion) constant, $y(k)$ is the (true baseband) equalizer output at baud instance k, and $E\{\cdot\}$ denotes expectation. The Constant Modulus Algorithm (CMA) performs a stochastic gradient descent of the CM cost function and updates the equalizer parameters according to the rule (see [6] or [12])

$$\mathbf{f}(k+1) = \mathbf{f}(k) + \mu \mathbf{r}^*(k) \underbrace{y(k)(\gamma^2 - |y(k)|^2)}_{CMA\ error\ term\ e_{bb}} \tag{2}$$

where $\mathbf{r}(k) = [\, r(k)\ \ r(k-1)\ \ldots\ r(k-M+1)\,]^T$ is a regressor vector of input samples to the equalizer tapped delay line, $\mathbf{f}(k)$ is the length-M vector of equalizer parameters, and $(\cdot)^*$ denotes conjugation. The term e_{bb} is referred to as the baseband CMA error term. We will drop the time index k in the sequel where risk of confusion is low.

When two-dimensional signaling is used (as in QAM) the equalizer output is complex and can be separated into in-phase (I) and quadrature (Q) components, $y = y_I + jy_Q$. The baseband CMA error term is decomposed into I and Q components as

$$e_{bb}^I(k) = y_I(k) \cdot (\gamma^2 - y_I(k)^2 - y_Q(k)^2) \tag{3}$$

$$e_{bb}^Q(k) = y_Q(k) \cdot (\gamma^2 - y_I(k)^2 - y_Q(k)^2) \tag{4}$$

Interestingly, CMA is globally convergent to an asymptotic setting which is equivalent within a phase shift to the Wiener setting (minimizing the desirable mean squared error (MSE) cost

ICASSP 2000 (TURKEY)

function) under a set of ideal conditions [3]. Even more remarkably, CMA exhibits extreme robustness properties to realistic signaling environments (summarized in [7]). Observe that the error calculation in (3) or (4) requires two (three) multiplies and one (two) add(s) for real (complex) signal processing. In high data rate scenarios, the computation of this error term and its application to the length$-M$ regressor vector can induce significant time delay which exceeds the symbol period. In this case, the current equalizer parameters are updated with an error term calculated using previous data, causing a performance degradation. Furthermore, multipliers can consume valuable chip area. We desire methods for quantizing the CMA error term which can be implemented using a look-up table so that chip area and signal latency can be reduced, and at the same time retain CMA's robustness properties and reduce the EMSE of CMA.

3. REAL-VALUED SIGNALING

Lucky's innovation [9] for LMS that replaces the equalizer output sample with its best estimate is called Decision-Directed LMS (DD-LMS). This idea can be extended to CMA and a Decision-Directed CMA (DD-CMA) algorithm is easily derived which retains CMA's robustness properties. The DD-CMA error term assumes a finite number of values and can be easily calculated with a look-up table. For the remainder of this section, all parameters are assumed to be real-valued.

Typically CMA is used as an acquisition algorithm, and adaptation is transferred to a Decision Directed (DD) mode of operation such as DD-LMS or a decision feedback equalizer (DFE) when error rates fall below a specified threshold. A decision device (sometimes called a "slicer") is therefore necessary hardware on the chip. A nearest-element decision device takes input x and selects the source alphabet member $\hat{x}$ with closest Euclidian distance to x as its output.

We propose to replace the usual equalizer output sample in the CMA error term with its best estimate from the decision device and name the algorithm Decision Directed CMA (DD-CMA). The baseband DD-CMA error term is therefore described by

$$\hat{e}_{bb} \quad = \quad \hat{y} \cdot (\gamma - \hat{y}^2) \tag{5}$$

where $\hat{y}$ is a best estimate of y. Observe that this function assumes a finite number of values since $\hat{y}$ assumes a finite number of values. Hence, the DD-CMA error term is easily calculated via a look-up table which is addressed using the current estimate $\hat{y}$. This look-up table replaces the two multiplies and one add needed to calculate the true CMA error term.

The solid line in Figure 1 is the true CMA error term in (2), and the dashed line is the Decision Directed CMA (DD-CMA) in (5) for 8-PAM signaling with symbol values $\{\pm 1, \pm 3, \pm 5, \pm 7\}$. The DD-CMA error term assumes six unique levels, with three unique magnitudes. Observe that for large-magnitude equalizer samples ($y > 6$), only one quantization level is assumed ($\hat{e}_{bb} = 84$). This coarse quantization in this area may slow initial algorithm convergence. Note also that the DD-CMA error term makes no sense for constant modulus source alphabets, since the error term is always zero.

3.1. Extension of Decision Device

By tolerating a small chip area increase, the quantization error variance can be substantially reduced by extending the slice levels

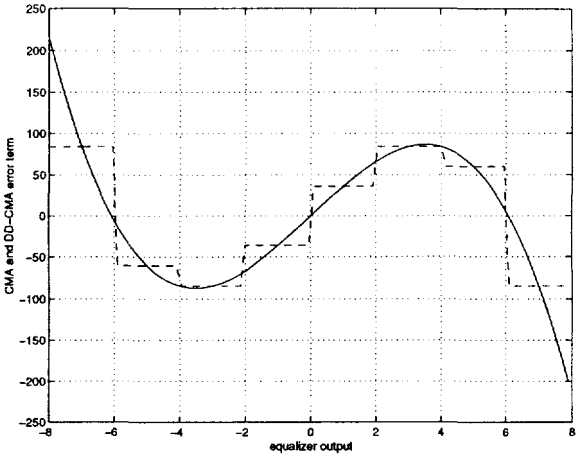


Figure 1: CMA (solid) and DD-CMA (dashed) error terms as a function of the equalizer output, y.

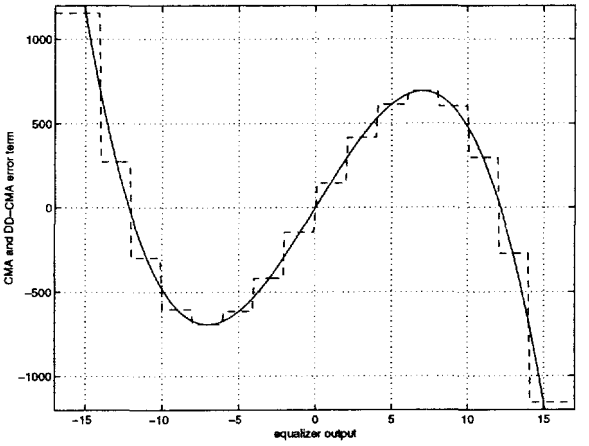


Figure 2: CMA (solid) and DD-CMA (dashed) error terms where the 8-PAM signal is quantized as a 16-PAM symbol.

of the decision device. For example, suppose the decision device is extended to slice 16-level as well as 8-level PAM signals. The 8-PAM symbol levels are denoted by $\{\pm 2, \pm 6, \pm 10, \pm 14\}$. The 16-level slice points are denoted by $\{\pm 1, \pm 3, \pm 5 \ldots \pm 13, \pm 15\}$. By quantizing the 8-PAM signal from the 16-PAM symbol set, the quantization error variance is reduced, as suggested in Figure 2, which shows the DD-CMA error term (dashed) and true CMA error term (solid).

Observe in Figure 2 that the DD-CMA error term assumes two levels ($\hat{e}_{bb} = 273, 1155$) for larger equalizer output values ($y > 12$), as opposed to one quantization level when an 8-level slicer is used (see Figure 1). Extension of the decision device to include 16-level over 8-level slicing can be accomplished with a minimal increase in combinatorial logic. The quantization error, however, is substantially improved.

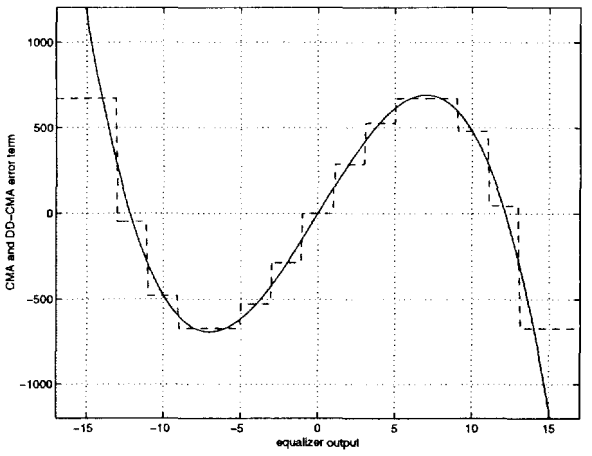


Figure 3: CMA (solid) and DD-CMA (dashed) where the slicer has been modified to yield mid-tread quantization.

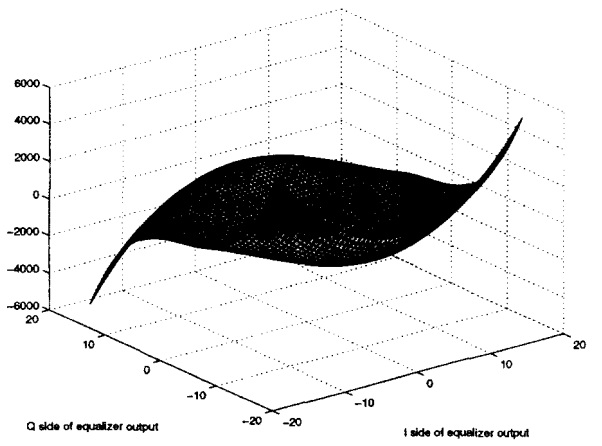


Figure 4: The in-phase component of the baseband CMA error term for 64-QAM signaling is illustrated as a surface above the $I - Q$ plane

3.2. Mid-Tread Quantization

Observe that the DD-CMA error term assumes non-zero values at the cubic root locations ($y = \{0, \pm\gamma\}$). This effect is in general termed mid-rise quantization, as compared to one which assumes zero values at the root locations and is denoted as mid-tread (see [8]). Another small modification to the decision device can be used to remove the bias from the slice values. For example, suppose the slice levels are chosen as $\{0, \pm 2, \pm 4, \ldots \pm 14\}$ instead of $\{\pm 1, \pm 3, \pm 5, \ldots \pm 15\}$. The 8-PAM symbol values are now equal to every other 16-level slice value. With decision boundaries chosen midway between slice points, the DD-CMA quantizes the CMA error term as shown in Figure 3 for 8-PAM signaling.

Mid-tread symbol quantization manifests as one true mid-tread and two near-mid-tread DD-CMA levels at the cubic roots. For example, the DD-CMA value is zero at $y = 0$, though it is (small but) non-zero near the roots at $\pm\gamma$. The DD-CMA value is not exactly zero near the $\pm\gamma$ roots since the source is multi-modulus. The Godard radius can be modified or the quantization value at the root locations can be manually set to zero to achieve true mid-tread quantization. In either case, since the zero-state is assumed at the cubic roots, we hypothesize that the excess mean-squared error (EMSE) (or stochastic jitter) of the DD-CMA update is reduced. (See [4] for the EMSE associated with the CMA update.)

4. COMPLEX-VALUED SIGNALING

The two-dimensional version of DD-CMA, illustrated with QAM signaling, is a simple extension of the one-dimensional case described above in §3. The baseband equalizer output (y), however, is now complex. Let y be represented by its real and imaginary parts, $y = y_I + jy_Q$. The baseband CMA error term can be written as

$$e_{bb} = (y_I + jy_Q) \cdot (\gamma^2 - y_I^2 - y_Q^2) \quad (6)$$

and separated into in-phase and quadrature components

$$e_{bb}^I = y_I \cdot (\gamma^2 - y_I^2 - y_Q^2) \quad (7)$$

$$e_{bb}^Q = y_Q \cdot (\gamma^2 - y_I^2 - y_Q^2) \quad (8)$$

Let $\hat{y}_I$ and $\hat{y}_Q$ be the output of the nearest-element decision device for inputs y_I and y_Q, respectively. The analogous DD-CMA to the real-signaling case is to replace y_I by $\hat{y}_I$ and y_Q by $\hat{y}_Q$ in each of the in-phase and quadrature components of the error term, or

$$\hat{e}_{bb}^I = \hat{y}_I \cdot (\gamma^2 - \hat{y}_I^2 - \hat{y}_Q^2) \quad (9)$$

$$\hat{e}_{bb}^Q = \hat{y}_Q \cdot (\gamma^2 - \hat{y}_I^2 - \hat{y}_Q^2) \quad (10)$$

Observe that the look-up table now requires a double indexing; $\hat{y}_I$ and $\hat{y}_Q$ must address the look-up table to calculate the baseband DD-CMA error term.

For 64-QAM signaling with I and Q symbol values chosen from $\{\pm 2, \pm 6, \pm 10, \pm 14\}$, each component of the CMA error term is a function of the two variables, y_I and y_Q. The in-phase component of the true CMA error term is plotted as a surface above the $I - Q$ plane in Figure 4.

Similarly, the in-phase component of the DD-CMA error term calculated according to (9) is plotted as a surface above the $I - Q$ plane in Figure 5, where the decision device is extended to 16 level slicing (see §3.1). The quantization of the error term over all $I - Q$ samples is quite good. We have found that extension of the decision device as done in this example can significantly improve algorithm performance, especially for lower-order QAM constellations.

5. PASSBAND IMPLEMENTATION

When the equalizer processes samples that are shifted away from DC, the equalizer is said to operate in the passband. Such a situation arises due to oscillator-drift in tuning, which can be tracked with a phase-locked loop. In this case, the baseband error term is rotated to form the passband error term, or $e_{pb} = e_{bb}e^{j\theta_k}$, where e_{pb} is the passband error term. We will describe the calculation of the passband DD-CMA error term for two-dimensional signaling.

The in-phase and quadrature components of the passband CMA error term are given by

$$e_{pb}^I = (\cos\theta_k y_I - \sin\theta_k y_Q) \cdot (\gamma^2 - y_I^2 - y_Q^2) \quad (11)$$

$$e_{pb}^Q = (\cos\theta_k y_Q + \sin\theta_k y_I) \cdot (\gamma^2 - y_I^2 - y_Q^2) \quad (12)$$

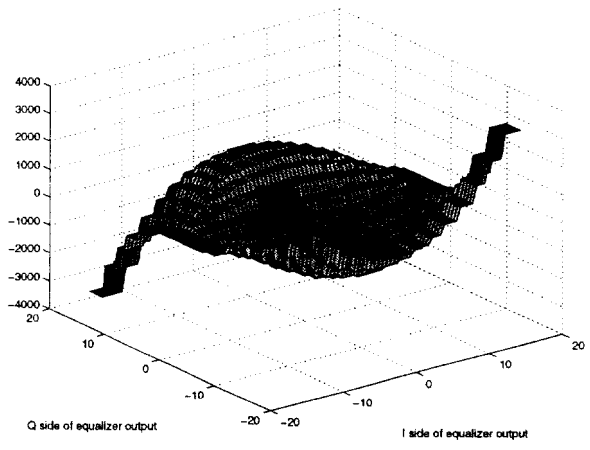


Figure 5: The in-phase component of the DD-CMA error term for 64-QAM signaling using a 16-level slicer is illustrated as a surface above the $I - Q$ plane.

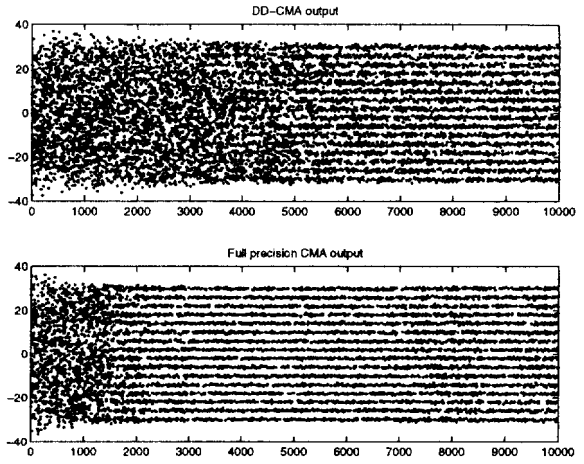


Figure 6: Equalizer ouput: (top) DD-CMA, (bottom) full precision CMA

for rotation angle θ_k.

Define the components of the passband equalizer output sample as $y_I^{pb} = \cos\theta_k y_I - \sin\theta_k y_Q$ and $y_Q^{pb} = \cos\theta_k y_Q + \sin\theta_k y_I$ and let their quantized values from the decision device be $\hat{y}_I^{pb}$ and $\hat{y}_Q^{pb}$, respectively. The passband DD-CMA update error term is found by replacing the baseband and passband samples in the passband CMA error term, (11) and 12), with their best estimates from the decision device,

$$\hat{e}_{pb}^I = \hat{y}_I^{pb} \cdot (\gamma^2 - \hat{y}_I^2 - \hat{y}_Q^2) \quad (13)$$

$$\hat{e}_{pb}^Q = \hat{y}_Q^{pb} \cdot (\gamma^2 - \hat{y}_I^2 - \hat{y}_Q^2) \quad (14)$$

The in-phase and quadrature components of the passband DD-CMA error term, $\hat{e}_{pb}^I$ and $\hat{e}_{pb}^I$, are each a function of three variables, the quantized baseband samples and one of the components of the quantized I and Q passband samples. Hence, the look-up table must be addressed by three values to calculate the passband DD-CMA error term.

6. SIMULATION EXAMPLE

Figure 6 shows the equalizer output for DD-CMA and full precision CMA using a 16 tap baud-spaced equalizer, and channel impulse response $[0.15, 0.93, -0.2, -0.06]$. The source alphabet is 16-PAM and DD-CMA uses a 32 level slicer (see §3.1). Notice that DD-CMA convergence is slower than full precision CMA.

7. CONCLUSION

This paper has presented novel techniques for calculating the update term for Godard's algorithm, or the Constant Modulus Algorithm (CMA). A look-up-table can be used in place or multipliers and adders to reduce chip area and signal latency. These methods work well and have been developed in silicon; see results of the NXT2000, a VSB/QAM demodulator, in [10].

8. REFERENCES

[1] D. R. Brown, P. Schniter, C. R. Johnson, Jr., "Computationally efficient blind equalization," *Allerton Conference on Communications, Control, and Computing*, Monticello, IL, pp. 54-63, Sept. 29 - Oct. 1, 1997.

[2] T. J. Endres, et al., "Adaptive equalizer with decision-directed constant modulus algorithm," *USPTO Patent pending*, s/n 09/261,843, filed March 3, 1999.

[3] I. Fijalkow, F. L. de Victoria, C. R. Johnson, Jr., "Adaptive fractionally-spaced blind equalization," *Signal Proc. Workshop on Stat. Signal and Array Proc.*, Yosemite Nat. Park, CA, pp. 257-60, Oct. 2-5, 1994.

[4] I. Fijalkow, C. E. Manlove, C. R. Johnson, Jr., "Adaptive fractionally-spaced blind CMA equalization: excess MSE," *IEEE Trans. on Signal Processing*, vol. 46, no. 1, pp. 227-231, Jan 1998.

[5] M. Ghosh, "A sign-error algorithm for blind equalization of real signals," *International Conf. on Acoustics, Speech and Signal Proc.*, Seattle, WA, pp. 3365-68, May 12-15, 1998.

[6] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Communications*, vol. 28, no. 11, pp. 1867-1875, Oct. 1980.

[7] C. R. Johnson, Jr., et al., "Blind equalization using the constant modulus criterion: a review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927-1950, Oct. 1998.

[8] N. S. Jayant, P. Noll, *Digital Coding of Waveforms*, Prentice Hall: Englewood Cliffs, NJ, 1984.

[9] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell Systems Technical Journal*, vol. 45, no. 2, pp. 255-286, Feb. 1966.

[10] P. Mannion, "VSB/QAM receiver cracks the code to indoor and mobile reception," *Electronic Design*, pp. 35-42, August 23, 1999.

[11] P. Schniter, C. R. Johnson, Jr., "Dithered signed-error CMA: robust, computationally efficient blind adaptive equalization," *IEEE Trans. on Signal Processing*, vol. 47, no. 6, pp. 1592-1603, June 1999.

[12] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. ASSP-31, no. 2, pp. 459-72, April 1983.

THE EFFECTS OF FINITE BIT PRECISION FOR A VLSI IMPLEMENTATION OF THE CONSTANT MODULUS ALGORITHM

L. R. Litwin, Jr.*, T. J. Endres†, S. N. Hulyalkar†, and M. D. Zoltowski*

Purdue University* and Sarnoff Digital Communications†

ABSTRACT

One of the most popular blind equalization techniques is the Constant Modulus Algorithm (CMA), and it has gained popularity in the literature and in practice because of its LMS-like complexity and its robustness to non-ideal, but practical, conditions. Although CMA has been well-studied in the literature, these analyses have typically implemented the algorithm using "infinite" precision arithmetic. The motivation for this paper is a VLSI implementation of a high data rate, fractionally spaced, linear forward equalizer whose taps are adjusted using CMA. In this paper we examine how implementing CMA using finite bit precisions affects the algorithm's performance.

1. INTRODUCTION

The Constant Modulus Algorithm (CMA) was originally proposed by Godard [1] for QAM signals and independently by Treichler and Agee [2] for constant envelope FM signals. CMA has seen recent theoretical growth (summarized in [3]) and has also been successfully deployed [4]. Although CMA has been deeply studied in the literature, virtually all of these analyses have been made under the assumption that the algorithm would be implemented using "infinite" precision. We will use the term infinite precision throughout this paper to refer to two cases: 1) when mathematics are performed under the assumption that each number has infinite precision, and 2) when simulations are run using numbers represented with the full precision of the computer. We are interested in the finite precision implementation of CMA and our motivation is a VLSI hardware implementation of CMA using fixed point arithmetic.

Section 2 is a brief primer to CMA and its tap update equation. Section 3 discusses the fixed point arithmetic that is typically used for high-speed hardware designs that involve mathematics. Section 4 describes the assumptions and settings used for the finite precision simulations of CMA. Section 5 presents the simulation results. Section 6 provides concluding remarks.

2. THE CONSTANT MODULUS ALGORITHM

CMA is a gradient descent technique that is used to minimize the cost function described by the Constant Modulus (CM) criterion. This cost function can be written as

$$J_{CM} = E\{(\gamma - |y_n|^2)^2\} \tag{1}$$

where γ is a positive constant known as the Godard radius and y_n represents the equalizer output at baud instance n.

The CMA tap update equation performs a stochastic gradient descent of J_{CM} and it is written as

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \mathbf{r}_n^* \underbrace{y_n(\gamma - |y_n|^2)}_{CMA\ error\ term} \tag{2}$$

where μ is a small, positive constant called the step-size, $\mathbf{f}_n$ is the length-M vector of equalizer taps at time n, and $\mathbf{r}_n = [\, r_n \;\; r_{n-1} \;\; r_{n-2} \ldots r_{n-M+1} \,]^T$ is the regressor vector formed from the M most recent equalizer input samples. Complex conjugation is denoted by the asterisk.

The successful use of CMA in practical receiver implementations has been a motivation for researching reduced complexity versions of CMA. A recent advance in the area of reduced complexity is signed-error CMA [5] which replaces the usual CMA error term with its signum (+/-1). An extension of signed-error CMA is dithered signed-error CMA [6] which introduces a dithering term to the CMA error term prior to the signum operation in order to preserve CMA's robustness properties.

Development of reduced complexity versions of CMA is important for high-speed VLSI applications because reduction in the complexity of the algorithm leads to hardware designs with higher operating speeds

*Department Of Electrical Engineering, Purdue University, West Lafayette, IN 47906, litwin(mikedz)@ecn.purdue.edu

†Sarnoff Digital Communications, Suite 100, 6 Penns Trail, Newtown, PA 18940, endres(samirh)@sdcomm.com

ICASSP 1999 (Phoenix, Az)

and/or lower gate counts. Hence, since (2) requires a full complex multiply for each equalizer tap, the implementation of (2) to process high data rate signals is often a computational bottleneck and the allocation of bit widths to the components of the CMA update equation is of foremost practical interest. This paper examines the effects of finite bit precision on the performance of CMA by assigning finite bit widths to the components of (2).

3. FIXED POINT ARITHMETIC

This section describes the fixed point representation of numbers and it also lists the major effects of using finite precision arithmetic.

3.1. Representation of Fixed Point Numbers

Although numbers in a VLSI design can be represented in floating point format (mantissa and exponent), fixed point arithmetic is the typical format for high speed designs. In particular, this paper focuses on the effects of implementing CMA using two's complement fixed point arithmetic.

In the two's complement fixed point format, a $B-$bit number's most significant bit (MSB) represents the sign of the number, and the lower $B-1$ bits represent the magnitude. Thus, a $B-$bit number can represent numbers from $-(2^{B-1})$ to $(2^{B-1}-1)$. In two's complement arithmetic, the negative of a binary number is formed by inverting each bit of the number and adding a 1 to the least significant bit (LSB).

3.2. Effects Of Finite Precision Arithmetic

Finite precision arithmetic can have severe effects on the algorithm's performance. In particular, there are three major types of errors that can occur due to finite precision arithmetic [7].

1. Data And Coefficient Quantization Errors

   When representing the data or the filter coefficients with B bits, the total number of possible values that the data or coefficients can take on is 2^B. Hence, the data or the coefficients are quantized to one of the 2^B levels, which introduces quantization noise resulting in a decrease in the effective SNR.

2. Truncation/Rounding Errors

   Truncation and rounding errors occur when a number is converted from a given precision to a lower precision. For example, when multiplying the data and the error term in (2), the bit width of the result increases (multiplying two numbers of bit widths a and b results in a number of bit width $a+b$). In order to maintain a reasonable complexity for the multipliers/adders, this result is truncated and errors are introduced due to the reduced precision.

3. Overflow/Underflow Errors

   Overflow (underflow) errors occur when the result of a calculation is too large (small) to be represented with the allocated bit width. When an arithmetic calculation results in an overflow (underflow), the result is typically clipped [7] to the maximum (minimum) value possible for that number of bits. This clipping operation induces nonlinear distortion in the signal.

The presence of these effects stresses the importance of assigning the appropriate number of bits to the components of each mathematical operation. The simulation results presented in this paper show how the manifestation of the above errors affects the performance of a finite precision VLSI implementation of CMA.

4. ASSUMPTIONS/PARAMETER SELECTIONS

This section describes the assumptions and parameter selections that were used for the simulation results presented in section 5.

4.1. Assumptions

All data is fractionally sampled with a spacing of $T/2$ where T is the symbol period. The channel models are derived from experimentally acquired microwave data and they are available at the SPIB database[1]. Specifically, channels 1 and 2 are used, and they are 300 and 230 samples long, respectively. The source sequence for each simulation consists of 150,000 $T/2$-spaced samples. The 75,000 symbols are taken from an equally probable 64-QAM alphabet. The mean-squared error (MSE) value that is shown in all of the figures is computed by averaging the instantaneous MSE over the last 25,000 symbols.

In order to simulate the fixed point arithmetic used in hardware, the result of each operation is rounded to an integer value and then clipped to make sure that the result is within the range of possible values for the number of assigned bits. The entire computation of the tap update term is performed with a precision of 32

[1]The Rice University Signal Processing Information Base (SPIB) channel database resides at http://spib.rice.edu/spib/microwave.html

bits. This high precision is used to isolate the effects of using reduced precisions for just the data and the taps.

We use B_{data} to denote the number of bits used to represent the data, and B_{taps} to denote the tap bit precision used when multiplying the data in the computation of (2). For the results shown in Figures 1 and 2, the taps are stored at 32-bit (full) precision, but only the upper B_{taps} bits are used to multiply the data. In Figure 3, the taps are stored at reduced precision, and only B_{taps} (where $B_{taps} < 32$) bits are used for both the storage of the tap values, and for the computation of (2).

4.2. Parameter Selections

1. A 16-tap fractionally-spaced equalizer (FSE) is used for Channel 1, and a 32-tap FSE is used for Channel 2.
2. A single spike initialization is used for the taps in all simulations. This initialization involves setting the center tap to unity and all other taps to zero. For the 16-tap equalizer, the center tap is tap position 8, and for the 32-tap equalizer, the center tap is tap position 16.
3. The value used for the step-size is $\mu = 2^{-22}$. This value is specifically chosen to be a power of two because it can be implemented in hardware as a simple right-shift by 22 bits as opposed to an actual multiply.

5. SIMULATION RESULTS

The approach taken in performing the simulations was to first hold the data precision at $B_{data} = 20$ bits while varying the tap precision from $B_{taps} = 4$ bits to 20 bits and recording the MSE for each setting. Subsequently, the tap precision was held at 20 bits while the data precision varied from 4 bits to 20 bits. No noise was present for either case. The purpose of this approach was to determine "good" settings to use for further simulations. The results of these simulations are shown in Figure 1. Note that a tap bit precision of 9 bits yields an MSE similar to that for a tap bit precison of 20 bits. Also note that a data bit precision of 6 bits yields an MSE similar to that for a data precision of 20 bits. The effects of quantization appear when using 5 bits for the data, although an aggressive design might perform adequately using this precision. However, 6 bits is preferable. 5 bits appears to be the lower limit, as evidenced by the jump of over 10 dB for the MSE when only 4 bits are used.

Although these simulations were only done for two channels, from the results we can extrapolate that, in general, the taps are more sensitive to finite bit precision effects compared to the data (this is an expected result based on the analysis of finite bit precision effects on FIR filters in general, for example, see [7]). Due to this increased sensitivity to finite bit precision, the taps of the CMA equalizer should be assigned a higher precision than that assigned to the data. In fact, based on the results shown here and on other simulation results, we hypothesize that a good rule of thumb for the tap bit precision is "data bits plus three," i.e. pick an acceptable value for B_{data} and assign $B_{taps} = B_{data} + 3$ bits for the taps. Based on this rule of thumb, we suggest that $B_{data} = 6$ and $B_{taps} = 9$ would be good precisions to use for 64-QAM and this class of channels, and for a more aggressive design, $B_{data} = 5$ deserves study.

In order to further test this hypothesis, we selected the more aggressive case of $B_{data} = 5$ and $B_{taps} = 9$ and ran simulations with an SNR level of 24 dB and for no noise. The results are presented in Figure 2. The top plot fixes $B_{data} = 5$ and varies the tap precision, while the bottom plot fixes $B_{taps} = 9$ and varies the data precision. Note that for our proposed precisions (and all precisions above that), the noise has the effect of "lifting" the curve. However, for the points below our proposed precisions, the effects of noise do not cause a significant change in performance. This is because the quantization effects dominate at these precisions. From these results, we can state that, for this class of channels, the values of $B_{data} = 5$ and $B_{taps} = 9$ are adequate precisions for 64-QAM, and one might want to use $B_{data} = 6$ to be conservative.

For the results presented above, the taps are stored at a precision of 32 bits, and only the upper B_{taps} are used to multiply the data. Further simulations were run to examine the effects of storing the taps at reduced precisions. The MSE trajectories are shown in Figure 3. In the case of reduced precision, the stated value for B_{taps} is the bit precision used for both storing the taps and used in (2). Note that when the taps are stored at the same reduced precision used to calculate (2), about twice the bit width is needed to achieve similar performance as when the taps are stored at higher precision. These results emphasize the importance of storing the taps at a high precision, even though only a lower precision is needed for the multiplication of the data.

6. CONCLUSIONS

We have shown how a fixed point VLSI implementation of CMA will suffer a performance loss due to the effects of finite bit precisions. Based on an admittedly limited data set, our results suggest that $B_{data} = 6$

bits and $B_{taps} = 9$ bits are good settings to use for 64-QAM, and $B_{data} = 5$ bits should be looked at for an aggressive design. Future work needs to look at how these settings perform on other channels. The results have also shown that the taps need to be stored at a higher precision than what is adequate for multiplying the data. Future work needs to determine a sufficient bit width for the tap storage. The CMA error term was computed using 32 bits, and more simulations need to be run to determine if a smaller bit width still provides adequate performance.

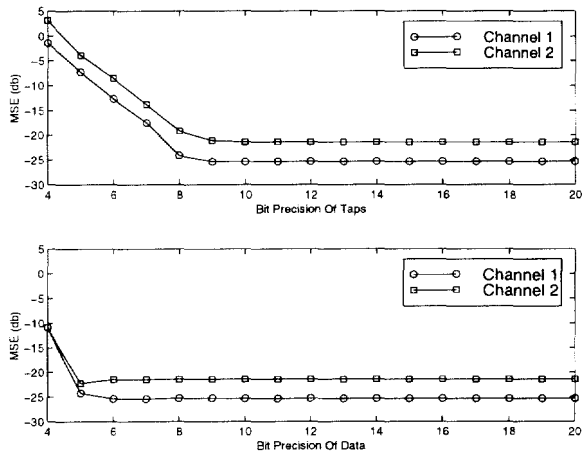


Figure 1: The top plot shows the MSE for $B_{data} = 20$ with varying tap precision and no noise. The bottom plot shows the MSE for $B_{taps} = 20$ with varying data precision and no noise.

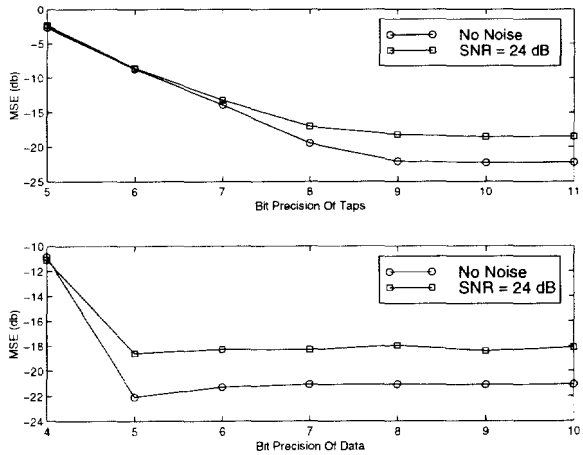


Figure 2: The top plot shows the MSE for $B_{data} = 5$ with varying tap precision. The bottom plot shows the MSE for $B_{taps} = 9$ with varying data precision. Both plots are for Channel 2.

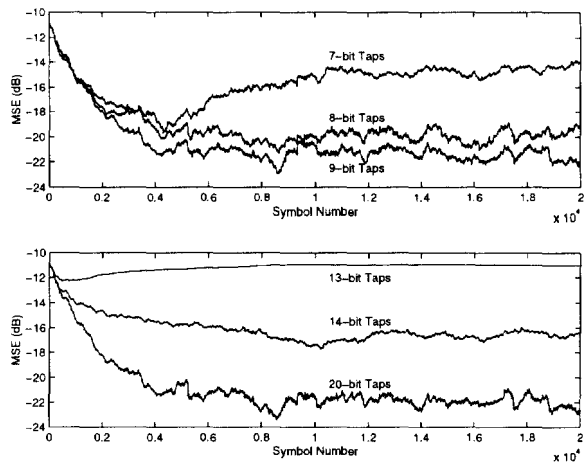


Figure 3: The top plot shows the MSEs that resulted from storing the taps at full (32-bit) precision. The bottom plot shows the MSEs that resulted from storing the taps at reduced (B_{taps}−bit) precision. Both plots are for Channel 2 with $B_{data} = 5$ and no noise.

7. REFERENCES

[1] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," IEEE Trans. on Communications, vol. 28, no. 11, pp. 1867-1875, Nov. 1980.

[2] J. R. Treichler, B. G. Agee, "A new approach to multipath correction of constant modulus signals," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-31, no. 2, pp. 459-72, Apr. 1983.

[3] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, R. A. Casas, "Blind Equalization Using The Constant Modulus Criterion: A Review," to appear in Proceedings of the IEEE, Sep. 1998.

[4] J. R. Treichler, M. G. Larimore, J. C. Harp, "Practical blind demodulators for high-order QAM signals," to appear in Proceedings of the IEEE, Sep. 1998.

[5] M. Ghosh, "A Sign-Error Algorithm For Blind Equalization Of Real Signals," International Conf. on Acoustics, Speech and Signal Proc., Seattle, WA, pp. 3365-68, May 12-15, 1998.

[6] P. Schniter, C. R. Johnson, Jr., "The Dithered Signed-Error Constant Modulus Algorithm," to appear in IEEE Transactions on Signal Processing, 1998.

[7] J. G. Proakis, D. G. Manolakis, Digital Signal Processing: Principles, Algorithms, and Applications, New Jersey: Prentice Hall, 1996.

PROCEEDINGS OF THE IEEE

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

OCTOBER 1998

Special Issue on:

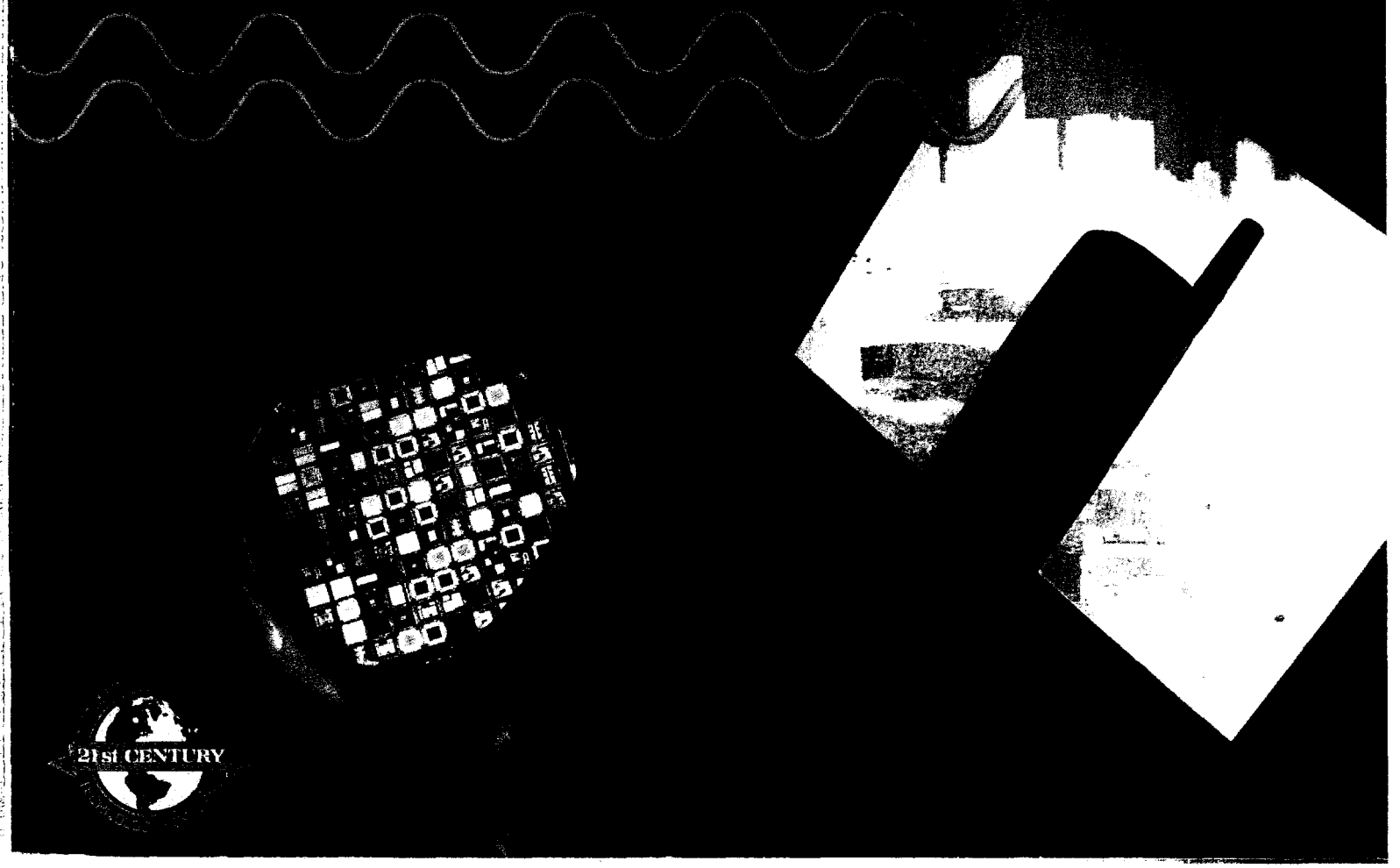
BLIND SYSTEM IDENTIFICATION AND ESTIMATION

Papers on: QAM Demodulators • CMA Criterion • Multichannel Id • Estimation Models • Beamforming Methods • Statistical Separation • Neural Network Approaches • Interference Suppression • Digital TV & Geophysical Applications

1928 Classic Paper: Radio Communications by Guglielmo Marconi

Predictive Paper: Some Thoughts on the State of the Technical Science in 2012 A.D.

Scanning the Past: The 75th Anniversary of the Naval Research Laboratory



Blind Equalization Using the Constant Modulus Criterion: A Review

C. RICHARD JOHNSON, JR., FELLOW, IEEE, PHILIP SCHNITER, THOMAS J. ENDRES, MEMBER, IEEE, JAMES D. BEHM, DONALD R. BROWN, AND RAÚL A. CASAS

Invited Paper

This paper provides a tutorial introduction to the constant modulus (CM) criterion for blind fractionally spaced equalizer (FSE) design via a (stochastic) gradient descent algorithm such as the constant modulus algorithm (CMA). The topical divisions utilized in this tutorial can be used to help catalog the emerging literature on the CM criterion and on the behavior of (stochastic) gradient descent algorithms used to minimize it.

Keywords— *Adaptive equalizers, blind deconvolution, blind equalization, constant modulus algorithm (CMA), digital communication, equalizers, intersymbol interference, least mean square methods.*

I. INTRODUCTION

Information-bearing signals transmitted between remote locations often encounter a signal-altering physical channel. Examples of common physical channels include coaxial, fiber optic, or twisted-pair cable in wired communications and the atmosphere or ocean in wireless communications. Each of these physical channels may cause signal distortion, including echoes and frequency-selective filtering of the transmitted signal. In digital communications, a critical manifestation of distortion is intersymbol interference (ISI), whereby symbols transmitted before and after a given symbol corrupt the detection of that symbol. All physical channels (at high enough data rates) tend to exhibit ISI. The presence of ISI is readily observable in the sampled impulse response of a channel; an impulse response corresponding to a lack of ISI contains a single spike of width less than the time between symbols. An example of a terrestrial microwave channel impulse response [obtained from the Rice University Signal Processing Information Base (SPIB)[1]] is shown in Fig. 1.

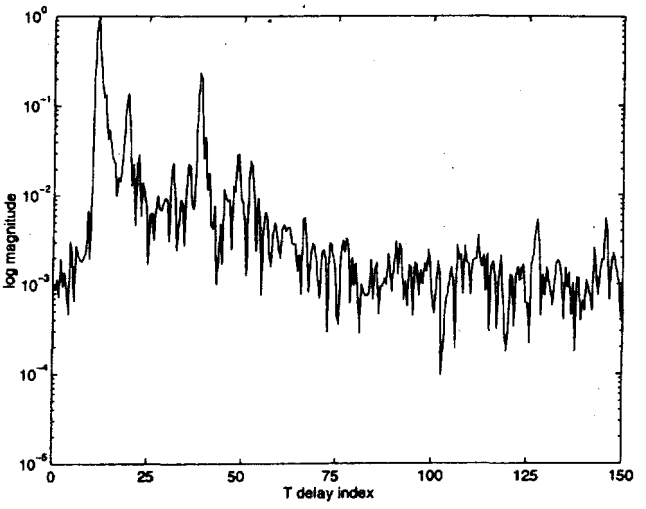


Fig. 1. Terrestrial microwave channel impulse response magnitude, $1/T = 30 \times 10^6$ symbols/s (SPIB channel #3).

Linear channel equalization, an approach commonly used to counter the effects of linear channel distortion, can be viewed as the application of a linear filter (i.e., the equalizer) to the received signal. The equalizer attempts to extract the transmitted symbol sequence by counteracting the effects of ISI, thereby improving the probability of correct symbol detection.

Since it is common for the channel characteristics to be unknown (e.g., at startup) or to change over time, the preferred embodiment of the equalizer is a structure adaptive in nature. Classical equalization techniques employ a time-slot (recurring periodically for time-varying situations) during which a training signal, known in advance by the receiver, is transmitted. The receiver adapts the equalizer (e.g., via LMS [6], [27]) so that its output closely matches the known

Manuscript received May 13, 1997; revised February 27, 1998. This work was supported in part by National Science Foundation Grant MIP-9509011 and Applied Signal Technology.

C. R. Johnson, Jr., P. Schniter, D. R. Brown, and R. A. Casas are with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: johnson@ee.cornell.edu; schniter@ee.cornell.edu; browndr@ee.cornell.edu; raulc@ee.cornell.edu).

T. J. Endres is with Sarnoff Digital Communications, Newtown, PA 18940 USA (e-mail: endres@erols.com).

J. D. Behm is with the Department of Defense, Ft. Meade, MD 20855 USA (e-mail: jimbehm@toad.net).

Publisher Item Identifier S 0018-9219(98)06972-2.

[1]This microwave channel database resides at http://spib.rice.edu/spib/microwave.html.

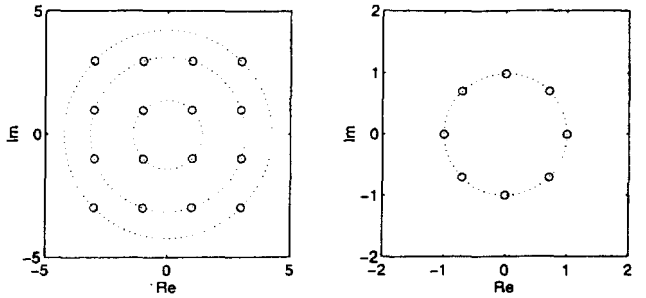


Fig. 2. (a) Nonconstant modulus source constellation (16-QAM) versus (b) CM source constellation (8-PSK).

reference training signal. Since the inclusion of such signals sacrifices valuable channel capacity, adaptation without resort to training, i.e., blind adaptation, is preferred. The most studied and implemented blind adaptation algorithm of the 1990's is the constant modulus algorithm (CMA).

CMA seeks to minimize a cost defined by the CM criterion. The CM criterion penalizes deviations in the modulus (i.e., magnitude) of the equalized signal away from a fixed value. In certain ideal conditions, minimizing the CM cost can be shown to result in perfect (zero-forcing) equalization of the received signal. Remarkably, the CM criterion can successfully equalize signals characterized by source alphabets not possessing a constant modulus [e.g., 16-quadrature amplitude modulation (QAM)], as well as those possessing a constant modulus (e.g., 8-PSK) (see Fig. 2). This paper attempts to explore the behavior of CMA by considering the similarities between the CM and mean-squared error (MSE) criteria. This relationship is important because of well-known connections between MSE and the actual quantity we desire to minimize: probability of bit error (e.g., see the discussion in [5]).

Plotting the CM cost versus the equalizer coefficients results in a surface referred to as the CM cost surface. Stochastic gradient descent (SGD) algorithms [6], [13] attempt to minimize the CM cost by starting at some location on the surface and following the trajectory of steepest descent. The CM cost surface characteristics are important because they can be used to understand the behavior of any SGD algorithms that attempt to minimize the CM cost, such as CMA. Specifically, these characteristics lend insight into the channel, equalizer, and source properties which affect SGD behavior.

The success of a stochastic gradient descent equalizer adaptation algorithm is dependent on a certain amount of stationarity in the received process. Thus, throughout the paper, we restrict our focus to stationary source and noise processes and to channels whose impulse response is fixed or slowly[2] time varying.

A. History

In the literature, blind equalization algorithms blossomed in the 1980's. The two principal precursors are Lucky's blind decision-direction algorithm [11] and Sato's algorithm [19]. What we term the CM criterion was introduced for blind equalization of QAM signals in [29] and of pulse-amplitude modulation (PAM) and FM signals in [30]. By the end of the 1980's blind equalizers were commercialized for microwave radio [9]. By the mid 1990's blind equalizers were realized in very large scale integration (VLSI) for high definition television (HDTV) set-top cable demodulators [23]. The current explosion of interest in the CM criterion stems from blind processing applications in emerging wireless communication technology (e.g., blind equalization, blind source separation, and blind antenna steering) and from CMA's record of practical success.

[2]Here "slow" is considered relative to the tracking speed of the SGD algorithm.

B. Our Mission

This paper is intended to be a resource to both readers experienced in blind equalization as well as those new to the subject. In a tutorial style, Section I-C provides background in fractionally spaced equalizer (FSE) modeling and design. (For baud-spaced equalizer (BSE) design, we refer the interested reader to a variety of classical references, e.g., [5], [6], [10], and [16].) Section II then illustrates several low-dimensional examples that help to characterize the behavior of FSE's adapted under the constant modulus criterion.

In Section III we construct a categorization of literature focusing on the application of the CM criterion to blind equalization. The annotated bibliography in Appendix III catalogs the existing literature according to the classifications of Section III, providing the reader with a valuable tool for further research. Our attempt to be exhaustive is justified only by the relative infancy of the subfield; evidence of the emerging status of this literature is seen in the wealth of conference papers in the bibliography of Appendix III.

Following the introductory FSE tutorial, Section I-E presents a novel view of classical nonblind adaptive equalization that illuminates the connection between the MSE and CM criteria. Specifically, the LMS-with-training strategy requires preselection of a design variable, namely training sequence delay, that may lead to a potentially suboptimal solution. The delay-optimized MSE (a function of equalizer parameters only) yields a cost surface (see Fig. 7) for which a simple LMS-like parameter update algorithm is not known to exist. Remarkably, the CM criterion offers a proxy for this surface for which there exists a (blind) parameter update algorithm, namely, CMA.

C. Fractionally Spaced Linear Equalization

In this section we describe the fractionally spaced equalization scenario and present some fundamental results regarding minimum MSE (i.e., Wiener [6]) equalizers. This material is primarily intended to provide background and context. For simplicity, our focus is restricted to a $T/2$-spaced FSE, where T denotes the baud, or symbol, duration. All results are extendible to the more general T/N-spaced case. Examples of seminal work on fractionally spaced

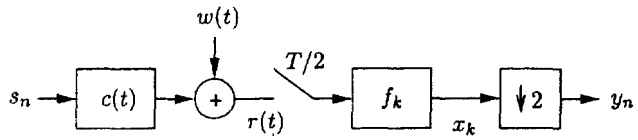


Fig. 3. Baseband model of single-channel communication system with $T/2$-spaced receiver.

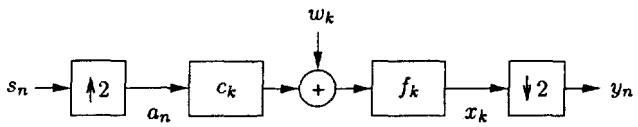


Fig. 4. Multirate system model.

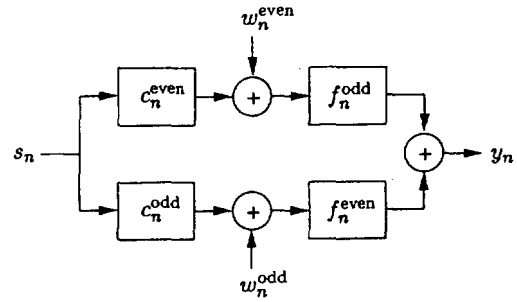


Fig. 5. Multichannel system model.

equalization include [4], [12], [14], and [26], while more comprehensive references are [5] and [18].

1) Multirate and Multichannel System Models: Consider the single-channel model illustrated in Fig. 3. A (possibly complex-valued) T-spaced symbol sequence $\{s_n\}$ is transmitted through a pulse shaping filter, modulated onto a propagation channel, and demodulated. We assume all processing between the transmitter and receiver is linear and time invariant (LTI) and can thus be described by the continuous-time impulse response $c(t)$. The received signal $r(t)$ is also corrupted by additive channel noise, whose baseband equivalent we denote by $w(t)$. The received signal is then sampled at $T/2$-spaced intervals and filtered by a $T/2$-spaced finite impulse response (FIR) equalizer of length $2N$. (An even length is chosen for notational simplicity.) This filtering can be regarded as a convolution of the sampled received sequence with the equalizer coefficients f_k. Finally, the FSE output $\{x_k\}$ is decimated by a factor of two to create the T-spaced output sequence $\{y_n\}$. Decimation is accomplished by disregarding alternate samples, thus producing the baud-spaced "soft decisions" y_n. We note that, in general, all quantities are complex valued. For clarity, we reserve the index n for baud-spaced quantities and the index k for fractionally spaced quantities throughout the paper.

Appendix I derives the equivalence between the continuous-time model in Fig. 3 and the discrete time models in Figs. 4 and 5, both constructed using $T/2$-spaced samples of $c(t)$ and $w(t)$. Fig. 4 depicts the multirate model while Fig. 5 depicts the multichannel model. Though our derivation of the discrete-time models is based on the single-channel system in Fig. 3, the equivalence between the multirate and multichannel models suggests that we could have based our model on a two-sensor (T-sampled) communication system instead. For a concise discussion on the equivalence between temporal and spatial diversity, see [15].

The multirate model of Fig. 4 uses the discrete-time fractionally spaced channel coefficients $c_k = c(k(T/2))$ and the discrete-time random process $w_k = w(k(T/2))$. The multichannel model of Fig. 5 subdivides these sample sequences into even and odd baud-spaced counterparts (of relative delay $T/2$), so that $c_n^{\text{even}} = c_{2n}$ and $c_n^{\text{odd}} = c_{2n+1}$ for $n = 0, 1, 2, \cdots$. In a similar manner, the FSE coefficients are partitioned as $f^{\text{even}} = f_{2n}$ and $f^{\text{odd}} = f_{2n+1}$.

Given a fractionally spaced channel of finite[3] length $2M$, we can collect the even and odd sets of equalizer and channel coefficients into column vectors

$$\begin{aligned}
\mathbf{f}_{\text{e}} &= [f_0, f_2, f_4, \cdots, f_{2N-2}]^t \\
&= \left[f_0^{\text{even}}, f_1^{\text{even}}, f_2^{\text{even}}, \cdots, f_{N-1}^{\text{even}}\right]^t \\
\mathbf{f}_{\text{o}} &= [f_1, f_3, f_5, \cdots, f_{2N-1}]^t \\
&= \left[f_0^{\text{odd}}, f_1^{\text{odd}}, f_2^{\text{odd}}, \cdots, f_{N-1}^{\text{odd}}\right]^t \\
\mathbf{c}_{\text{e}} &= [c_0, c_2, c_4, \cdots, c_{2M-2}]^t \\
&= \left[c_0^{\text{even}}, c_1^{\text{even}}, c_2^{\text{even}}, \cdots, c_{M-1}^{\text{even}}\right]^t \\
\mathbf{c}_{\text{o}} &= [c_1, c_3, c_5, \cdots, c_{2M-1}]^t \\
&= \left[c_0^{\text{odd}}, c_1^{\text{odd}}, c_2^{\text{odd}}, \cdots, c_{M-1}^{\text{odd}}\right]^t. \qquad (1)
\end{aligned}$$

It is possible to form the (baud-spaced) impulse response of the linear system relating s_n to y_n using a pair of $P \times N$ baud-spaced convolution matrices $\mathbf{C}_{\text{e}}$ and $\mathbf{C}_{\text{o}}$, where $P = M + N - 1$.

$$\mathbf{C}_{\text{e}} = \begin{bmatrix} c_0^{\text{even}} & & & \\ c_1^{\text{even}} & c_0^{\text{even}} & & \\ \vdots & c_1^{\text{even}} & & \\ c_{M-1}^{\text{even}} & \vdots & \ddots & c_0^{\text{even}} \\ & c_{M-1}^{\text{even}} & & c_1^{\text{even}} \\ & & & \vdots \\ & & & c_{M-1}^{\text{even}} \end{bmatrix}$$

$$\mathbf{C}_{\text{o}} = \begin{bmatrix} c_0^{\text{odd}} & & & \\ c_1^{\text{odd}} & c_0^{\text{odd}} & & \\ \vdots & c_1^{\text{odd}} & & \\ c_{M-1}^{\text{odd}} & \vdots & \ddots & c_0^{\text{odd}} \\ & c_{M-1}^{\text{odd}} & & c_1^{\text{odd}} \\ & & & \vdots \\ & & & c_{M-1}^{\text{odd}} \end{bmatrix}. \qquad (2)$$

[3] In practice, we would consider the fractionally spaced channel to be of "finite length" M if the response magnitude can be said to decay below some sufficiently small threshold for all time $t \geq M(T/2)$.

Convolution matrices are constructed so that, for example, the vector $\mathbf{C}_o\mathbf{f}_e$ is composed of coefficients from the convolution $c_n^{\text{odd}} \star f_n^{\text{even}}$.

Defining the compound matrix and vector quantities

$$\mathbf{C} = [\mathbf{C}_o \quad \mathbf{C}_e] \qquad \mathbf{f} = \begin{bmatrix} \mathbf{f}_e \\ \mathbf{f}_o \end{bmatrix} \tag{3}$$

we can rewrite the noise-free multichannel convolution equation (39) compactly in terms of the P (baud-spaced) system impulse response coefficients $\mathbf{h} = [h_0, h_1, \cdots, h_{P-1}]^t$

$$\mathbf{h} = \mathbf{C}\mathbf{f}. \tag{4}$$

Equation (4) indicates that $\mathbf{C}$ maps the FSE coefficient vector to its corresponding system response. Note that $\mathbf{C}$ is a Sylvester matrix [8].

A different (though essentially equivalent) construction of $\mathbf{C}$ and $\mathbf{f}$ deserves mention. First consider the fractionally spaced convolution matrix $\mathbf{C}_{FS}$ constructed as in either $\mathbf{C}_e$ or $\mathbf{C}_o$ in (2), but from a vector of fractionally spaced channel coefficients $\mathbf{c}_{FS} = [c_0, c_1, c_2, \cdots, c_{2M-1}]^t$

$$\mathbf{C}_{FS} = \begin{bmatrix} c_0 & & & \\ c_1 & c_0 & & \\ c_2 & c_1 & \ddots & \\ \vdots & c_2 & \ddots & c_0 \\ c_{2M-1} & \vdots & \ddots & c_1 \\ & c_{2M-1} & & c_2 \\ & & \ddots & \vdots \\ & & & c_{2M-1} \end{bmatrix}. \tag{5}$$

The product of $\mathbf{C}_{FS}$ with FSE coefficient vector $\overline{\mathbf{f}} = [f_0, f_1, f_2, \cdots, f_{2N-1}]^t$ yields the fractionally spaced impulse response between the upsampler and downsampler in Fig. 4, i.e., $\mathbf{h}_{FS} = \mathbf{C}_{FS}\overline{\mathbf{f}}$ [See (47) in Appendix I-B.] Since the baud-spaced impulse response $\mathbf{h}$ is formed using the odd coefficients of $\mathbf{h}_{FS}$, we reason that $\mathbf{h}$ can be constructed from the product of $\overline{\mathbf{f}}$ and a row-decimated version of $\mathbf{C}_{FS}$. In other words, $\mathbf{h} = \overline{\mathbf{C}}\overline{\mathbf{f}}$ where $\overline{\mathbf{C}}$ is formed from the odd[4] rows of $\mathbf{C}_{FS}$

$$\overline{\mathbf{C}} = \begin{bmatrix} c_1 & c_0 & & & & \\ c_3 & c_2 & c_1 & c_0 & & \\ \vdots & \vdots & c_3 & c_2 & \ddots & \\ c_{2M-1} & c_{2M-2} & \vdots & \vdots & \ddots & c_0 \\ & & c_{2M-1} & c_{2M-2} & & c_2 \\ & & & & \ddots & \vdots \\ & & & & & c_{2M-2} \end{bmatrix}. \tag{6}$$

Notice that $\overline{\mathbf{C}}$ is a column reordering of $\mathbf{C}$ and $\overline{\mathbf{f}}$ is a row reordering of $\mathbf{f}$. Thus, we consider the alternate formulation of the "decimated fractionally spaced convolution matrix" $\overline{\mathbf{C}}$ in (6) as essentially equivalent to $\mathbf{C}$ in (3).

[4]Throughout, we assume a vector/matrix indexing that starts with zero rather than one, so that the first row is considered "even" and the second "odd."

The convention we adopt in constructing $\mathbf{C}$ and $\overline{\mathbf{C}}$, which is sometimes referred to as "odd-sampled" decimation, connects the odd subchannel output to the even subequalizer input and vice versa (see Fig. 5). Appendix I discusses the implications of this choice.

In the baud-spaced equalization context [10], [16], the convolution matrix $\mathbf{C}_{BS}$ relating the equalizer coefficient vector to the baud-spaced impulse response does not have the compound form of (3) or (6). Instead it appears like $\mathbf{C}_e$ (or $\mathbf{C}_o$) in (2), but with columns constructed from the T-spaced samples of the channel response. In the absence of channel noise, this construction of $\mathbf{C}_{BS}$ yields the BSE design equation

$$\mathbf{h} = \mathbf{C}_{BS}\mathbf{f}_{BS} \tag{7}$$

where $\mathbf{f}_{BS}$ is the baud-spaced equalizer coefficient vector.

2) Requirements for Perfect Source Recovery: Equation (4) leads to what are commonly referred to as the "length and zero" conditions for perfect fractionally spaced equalization. We use the term perfect equalization interchangeably with perfect source recovery (PSR), i.e., $y_n = s_{n-\delta}$ for some fixed delay δ and any source sequence $\{s_n\}$. In addition to the absence of noise, PSR requires the "zero-forcing" system impulse response

$$\mathbf{h}_\delta = [0 \cdots 0, \ 1, \ 0 \cdots 0]^t \tag{8}$$

where the nonzero coefficient is in the δth position (and δ must satisfy $0 \le \delta \le P-1$). This response characterizes a system which merely delays the transmitted symbols by δ baud intervals. In order to achieve this particular response, the system of linear equations described by $\mathbf{h}_\delta = \mathbf{C}\mathbf{f}$ must have a solution. For PSR under arbitrary δ,[5] $\mathbf{C}$ must be full row rank [22]. This condition is sometimes referred to as strong perfect equalization.

The full-rank requirement implies that $\mathbf{C}$ must have at least as many columns as rows, which, in the $T/2$-spaced case, results in the following equalizer length requirement:

$$2N \ge M + (N-1) \Rightarrow N \ge M-1. \tag{9}$$

Applying the same argument to (7) reveals the reason that no FIR BSE can perfectly equalize a nontrivial FIR channel: the row dimension of $\mathbf{C}_{BS}$ always exceeds the column dimension. The $T/2$-spaced full rank requirement also implies that the polynomials specified by the coefficients $\mathbf{c}_e$ and $\mathbf{c}_o$ share no common roots (i.e., the polynomials are coprime). Appendix I-C discusses this common-root condition in more detail.

D. Mean-Square Error Criterion

In the presence of noise, we desire to minimize the expected squared magnitude of the recovery error

$$e_n = y_n - s_{n-\delta} \tag{10}$$

[5]A necessary and sufficient condition on perfect equalization (in the absence of noise) is that there exist a δ for which $\mathbf{h}_\delta$ lies in the column space of $\mathbf{C}$. Hence, there exist channels that do not result in full row-rank convolution matrices but that do satisfy $\mathbf{h}_\delta = \mathbf{C}\mathbf{f}$ for particular δ. Though we acknowledge the existence of such channels, we consider them to be trivial in the physical sense.

for a particular choice of delay (δ). We will see that this criterion can be interpreted as the best compromise between ISI and noise amplification in a minimum mean-squared error (MMSE) sense.

To formulate this error criterion more precisely, we collect the P previous T-spaced elements of the source sequence into the vector

$$\mathbf{s}(n) = \left[s_n, s_{n-1}, s_{n-2}, \cdots, s_{n-(P-1)}\right]^t \quad (11)$$

and the last $2N$ fractionally sampled values of noise into vector $\mathbf{w}(n)$

$$\mathbf{w}(n) = \left[w_{n-1}, w_{n-3}, w_{n-5}, \cdots, w_{n-(2N-1)}, w_n, w_{n-2}, w_{n-4}, \cdots, w_{n-(2N-2)}\right]^t \quad (12)$$

where the collection of even noise samples follows the collection of odd noise samples to be consistent with our definitions of $\mathbf{C}$ and $\mathbf{f}$ in (3). [Note, however, that this particular ordering of samples in the noise vector is inconsequential when assuming an independent identically distributed (i.i.d.) noise process.] With these quantities, the nth equalizer output $y_n = y(nT + (T/2))$ can be written compactly as

$$y_n = \mathbf{s}^t(n)\mathbf{C}\mathbf{f} + \mathbf{w}^t(n)\mathbf{f} \quad (13)$$

yielding an expression for the recovery error

$$e_n = \mathbf{s}^t(n)(\mathbf{C}\mathbf{f} - \mathbf{h}_\delta) + \mathbf{w}^t(n)\mathbf{f}. \quad (14)$$

Under the assumption that the noise and source processes are i.i.d. and jointly uncorrelated, with respective variances σ_w^2 and σ_s^2, the expected value of the magnitude-squared recovery error becomes

$$E\{|e_n|^2\} = (\mathbf{C}\mathbf{f} - \mathbf{h}_\delta)^H(\mathbf{C}\mathbf{f} - \mathbf{h}_\delta)\sigma_s^2 + \mathbf{f}^H\mathbf{f}\sigma_w^2. \quad (15)$$

(Appendix I-D discusses the independence assumption regarding fractionally sampled channel noise.) Note that (15) is proportional to the source-power-normalized MSE cost function

$$J_{\mathrm{MSE}} = (\mathbf{C}\mathbf{f} - \mathbf{h}_\delta)^H(\mathbf{C}\mathbf{f} - \mathbf{h}_\delta) + \lambda\mathbf{f}^H\mathbf{f} \quad (16)$$

where $\lambda = \sigma_w^2/\sigma_s^2$. In terms of $\mathbf{A} = \mathbf{C}^H\mathbf{C} + \lambda\mathbf{I}$, the technique of "completing the square" yields

$$J_{\mathrm{MSE}} = \left(\mathbf{f} - \mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta\right)^H\mathbf{A}\left(\mathbf{f} - \mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta\right) - \mathbf{h}_\delta^H\mathbf{C}\mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta + \mathbf{h}_\delta^H\mathbf{h}_\delta. \quad (17)$$

Note that $\mathbf{A}$ is positive definite for $\lambda > 0$.

Equation (17) indicates that the equalizer parameter vector minimizing J_{MSE} is

$$\mathbf{f}^\dagger = \mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta \quad (18)$$

and it follows that the $\mathbf{f}$-optimal MSE

$$\min_{\mathbf{f}} J_{\mathrm{MSE}}(\mathbf{f}, \delta) = J_{\mathrm{MSE}}(\mathbf{f}^\dagger, \delta) \quad (19)$$

$$= \mathbf{h}_\delta^H\left(\mathbf{I} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^H\right)\mathbf{h}_\delta \quad (20)$$

remains a function of system delay δ. We make this property explicit by adopting the notation $J_{\mathrm{MSE}}(\mathbf{f}, \delta)$. It follows from (8) and (20) that the optimum delay $\delta^\dagger$ corresponds to the index of the minimum diagonal element of $\mathbf{I} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^H$ [7]. This is written formally as

$$\delta^\dagger = \arg\min_\delta \left\{\left[\mathbf{I} - \mathbf{C}\left(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I}\right)^{-1}\mathbf{C}^H\right]_{\delta,\delta}\right\}. \quad (21)$$

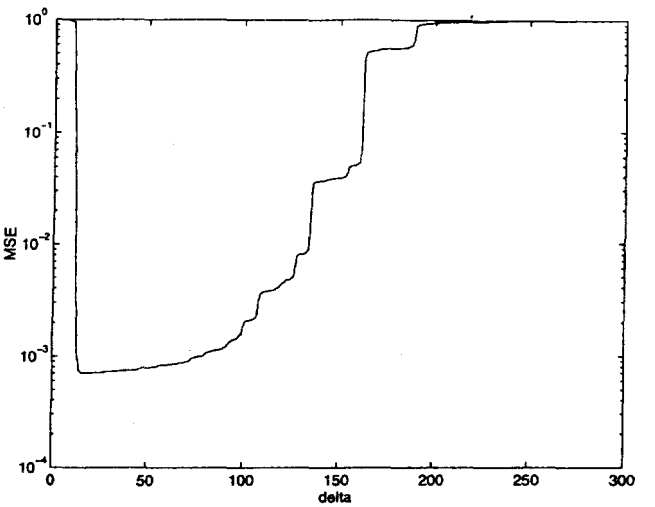


Fig. 6. $\mathbf{f}$-optimal MSE, $J_{\mathrm{MSE}}(\mathbf{f}^\dagger, \delta)$, versus T-spaced delay δ for the channel of Fig. 1 and 30 dB SNR using 300-tap FSE.

For a $T/2$-spaced FSE with 300 taps and an SNR ($= 1/\lambda$) of 30 dB, Fig. 6 plots $J_{\mathrm{MSE}}(\mathbf{f}^\dagger, \delta)$ versus δ for the "typical" impulse response of Fig. 1. Note the degree to which δ can affect MSE performance.

We conclude that proper preselection of δ is important for equalizer-based minimization of $J_{\mathrm{MSE}}(\mathbf{f}, \delta)$. This idea of fixed-$\delta$ optimization is of particular relevance because it describes the typical adaptive equalization scenario when a training signal is available [17].

E. An Amalgamated MSE Cost Function

When the source is differentially encoded [5], knowledge of absolute phase is not required for symbol detection. For example, either $y_n = s_{n-\delta}$ or $y_n = -s_{n-\delta}$ (for all n) would form an acceptable output sequence for differentially encoded binary phase-shift keying (BPSK). (For complex-valued source alphabets such as QAM, we allow $y_n = e^{j(\pi/2)m}s_{n-\delta}$ for fixed $m \in \{0, 1, 2, 3\}$.) Therefore, an acceptable system impulse response can include a fixed phase shift in addition to a bulk system delay δ. With this in mind, we construct a phase- and delay-optimized amalgamated cost function $J_{\mathrm{A}}(\mathbf{f})$

$$J_{\mathrm{A}}(\mathbf{f}) = \min_{\delta,\rho}\left\{(\mathbf{C}\mathbf{f} - \rho\mathbf{h}_\delta)^H(\mathbf{C}\mathbf{f} - \rho\mathbf{h}_\delta) + \lambda\mathbf{f}^H\mathbf{f}\right\} \quad (22)$$

where ρ is one of the set of allowable phase shifts (e.g., $\{+1, -1\}$ for real-valued PAM).

J_{A} is a multimodal fabrication, bearing similarity to a $(2N+1)$-dimensional egg carton. A surface plot appears in Fig. 7 for well-behaved $T/2$-spaced channel $\mathbf{c}_1$ defined in Table 1. By "well-behaved" we mean that $\mathbf{c}_1$ has no common or nearly common subchannel roots. Fig. 7 indicates that if we minimize $J_{\mathrm{A}}(\mathbf{f})$ by a gradient descent strategy, then the initial value of $\mathbf{f}$ will determine the values

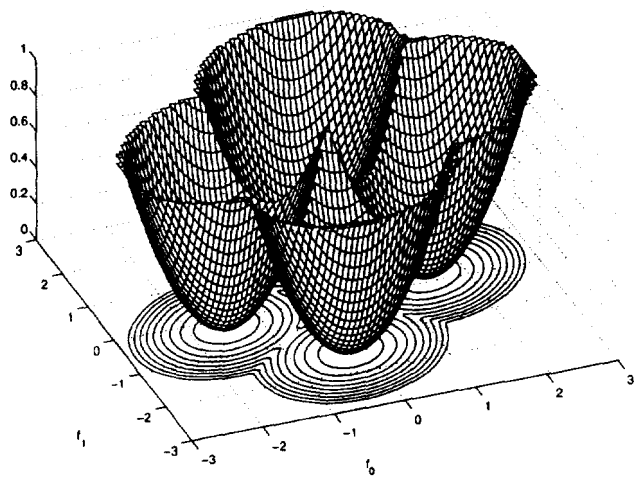


Fig. 7. J_{A} for well-behaved channel $\mathbf{c}_1$ and no noise in equalizer ($\mathbf{f}$) space.

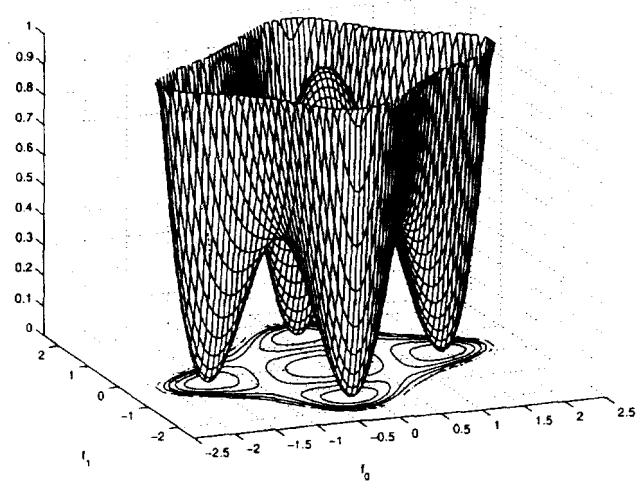


Fig. 8. J_{CM} for well-behaved channel $\mathbf{c}_1$ and no noise in equalizer ($\mathbf{f}$) space.

Table 1 Summary of Channels Used for Two-Tap FSE Examples

Name	$T/2$-spaced Impulse Response	Classification
$\mathbf{c}_1$	[-0.0901, 0.6853, 0.7170, -0.0901]	well-behaved[7]
$\mathbf{c}_2$	[1.0, -0.5, 0.2, 0.3]	well-behaved
$\mathbf{c}_3$	[-0.0086, 0.0101, 0.9999, -0.0086]	nearly-common subchannel roots
$\mathbf{c}_4$	[1.0, -0.5, 0.2, 0.3, -0.2, -0.15]	undermodelled

of δ and ρ to which the descent scheme will asymptotically converge. In other words, optimization of $J_{\mathrm{A}}(\mathbf{f})$ by gradient descent accomplishes preselection of δ via choice of $\mathbf{f}$-initialization.

Section II attests to the claim that *the CM criterion serves as a close proxy to J_{A}, which is robust under typical operating conditions.* For a preview, compare the CM cost surface in Fig. 8 to the amalgamated MSE surface in Fig. 7 for the same channel $\mathbf{c}_1$. As such, the CM criterion offers a performance metric that bears many similarities to MSE but which is capable of minimization by (stochastic) gradient descent schemes conducted blindly with respect to the transmitted symbols.

With our tutorial orientation, Section II restricts focus to a two-tap FSE design task that permits visualization of equalizer-parameter-space cost-contour plots illustrating various properties of the CM cost function J_{CM}. In particular, we can isolate an "ideal zero-cost" situation where the stationary points in J_{CM} and J_{A} match exactly and where the minima achieve zero cost. This special case requires several assumptions not often satisfied in practice. We will examine examples of CM-adapted FSE behavior conducted under violations of these requirements for ideal zero-cost equalization. This implicit taxonomy will be used in Section III to provide an overview of the literature citations in the annotated bibliography of Appendix III.

II. Two-Tap Illustrative Examples

The shape of the cost surface defining a particular stochastic gradient algorithm often lends great insight into the expected behavior of that algorithm. With this in mind, we embark on a tutorial study of the cost surface defined by the CM criterion and descended by CMA. First, however, consider the following list of features characterizing a generic (stochastic) gradient descent algorithm.

- — Far from a stationary point, the gradient (i.e., first derivative) of the cost surface determines local convergence rate.
- — Near a stationary point, the local curvature (i.e., second derivative) of the cost surface determines local convergence rate.
- — Local minima with nonzero cost induce excess steady-state error in stochastic gradient descent algorithms with nonvanishing step-sizes.
- — Multimodal surfaces may exhibit local minima of varying cost, thus linking initialization to achievable asymptotic performance.
- — "Poor" initialization on a multimodal surface can lead a trajectory into temporary capture by (one or more) saddle points, resulting in arbitrarily slow convergence to a minimum.
- — Nontrivial deformations of a multimodal surface relocate each saddle point and alter the region of attraction associated with each local minima.

The following sections combine low-dimensional examples with the well-known characteristics above to formulate an intuitive understanding of the CM criterion and its connection to the MSE criterion.

A. Two-Tap Equalizer Design Equations

As discussed in Section I-C, satisfaction of the "length and zero" conditions ensures an exact solution to the zero-forcing equation $\mathbf{h}_\delta = \mathbf{C}\mathbf{f}$. For a two-tap $T/2$-spaced FSE, the length condition is satisfied for channels with impulse responses $[c_0, c_1, c_2, c_3]$ and shorter. For a length-four channel, the root condition is satisfied when the even and odd subchannel polynomials $C_{\mathrm{even}}(z^{-1}) = c_0 + c_2 z^{-1}$ and $C_{\mathrm{odd}}(z^{-1}) = c_1 + c_3 z^{-1}$ have distinct roots.

In this case, (3) specifies that the FSE design quantities take the following form:

$$\mathbf{C} = \begin{bmatrix} c_1 & c_0 \\ c_3 & c_2 \end{bmatrix} \qquad \mathbf{f} = \begin{bmatrix} f_0 \\ f_1 \end{bmatrix}. \tag{23}$$

Since $\mathbf{h}_\delta$ has one nonzero coefficient, the zero-forcing equalizer will be proportional to either the first or the second column of $\mathbf{C}^{-1}$. Thus, all four channel parameters enter into the design of $\mathbf{f}$; the sub-equalizers of Fig. 5 are not simply inverses of their respective subchannels.

B. Introduction to the CM Cost Function

The CM cost function can be motivated using the temporary assumption that the source is binary valued (± 1). In this case, s_n has a constant squared-modulus of one ($|s_n|^2 = 1$). Under perfect symbol recovery we know that the output y_n has the same CM property and can thus imagine a cost that penalizes deviations from this output condition. This, in fact, defines the CM cost function for a BPSK source

$$J_{\mathrm{CM}}|_{\mathrm{BPSK}} = E\left\{\left(1 - |y_n|^2\right)^2\right\}.$$

Appendix II presents more general versions of the CM cost function and derives expressions for J_{CM} in terms of channel parameters, particular source and noise statistics, and equalizer coefficients.

The leap of faith, first espoused by [29], is the application of J_{CM} to a multilevel (i.e., nonconstant modulus) source. Reference [29], which addressed baud-spaced blind equalization via minimization of J_{CM}, makes the first observation concerning the proximity of the J_{CM} and J_{A} minima.

> It should also be noted that the equalizer coefficients minimizing the dispersion functions closely approximate those which minimize the mean squared error.

This is remarkable because an approximation of J_{CM} can be formed solely from the equalizer output y_n; no training signal is required to compose an accurate gradient approximation for use in a stochastic gradient minimization algorithm such as CMA [30]. It is worth noting that the phase-independent nature of J_{CM} has its own advantages in modem design [24].

C. Illustrative Cost Surface Examples

The following sections present mesh and contour plots of the CM cost surface for a two-tap $T/2$-spaced FSE under various operating conditions. Refer to Table 1 for definitions of the various channels used in our experiments. In all contour plots, the asterisks ($*$) indicate the locations of global MSE (i.e., J_{A}) minima while the crosses ($\times$) indicate the locations of local MSE minima. Recall that different pairs of MSE minima (reflected through the origin) correspond to different values of system delay, while the two elements composing each pair correspond to the two

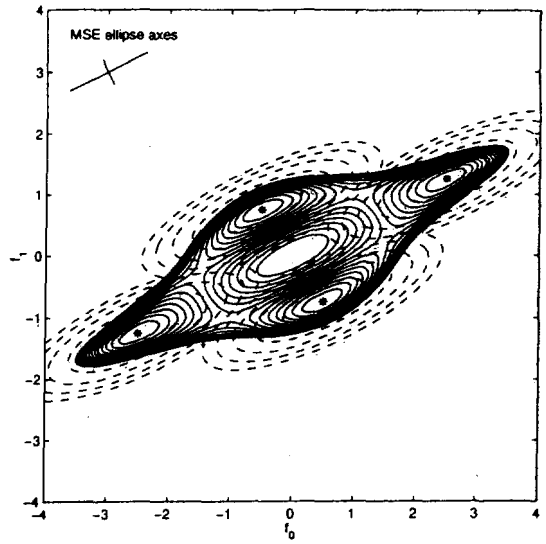


Fig. 9. J_{CM} contours (solid) for well-behaved channel $\mathbf{c}_2$ and no noise, with J_{A} overlay (dashed) and global MSE minima marked by "$*$" in equalizer ($\mathbf{f}$) space.

choices of system polarity.[6] Thus, the asterisks mark the MMSE equalizers of optimum system delay. The "MSE ellipse axes" appearing in the upper left corner of each contour plot indicate the orientation and eccentricity of the elliptical MSE contours (see Fig. 9).

All quantities in the experiments are real-valued. Unless otherwise noted, the source used was zero-mean and i.i.d. with alphabet $\{-1, ,1\}$.

1) Ideal Zero-Cost Equalization: For a well-behaved[7] channel $\mathbf{c}_1$ in the absence of channel noise, Fig. 8 plots J_{CM} in equalizer space. Recall that Fig. 7 plots J_{A} for the same noiseless channel. For a different well-behaved and noiseless channel $\mathbf{c}_2$, Fig. 9 superimposes the corresponding J_{CM} and J_{A} cost contours. Note the symmetry (with respect to the origin) exhibited by both J_{CM} and J_{A} cost surfaces.

In these ideal situations, all MSE and CM minima attain costs of zero (see Figs. 7 and 8). In addition, it can be seen that the locations of the J_{CM} and J_{A} minima coincide. (The J_{CM} minima locations can be inferred from the J_{CM} cost contours.) Fig. 9 also indicates that the curvatures of CM and MSE cost surfaces in the neighborhoods of local minima are closely related.

2) Combined Channel-Equalizer Space: The behavior of a gradient descent of J_{CM} is sometimes studied in the (downsampled) combined channel-equalizer space (i.e., $\mathbf{h}$ from Section I-C). The appeal of studying J_{CM} in $\mathbf{h}$-space follows from the normalization and alignment of J_{CM} with the coordinate axes. These features are clear in a comparison of Figs. 9 and 10, both constructed from the same noiseless channel. Equation (4) implies that a unique reversible mapping (i.e., an isomorphism) exists between

[6] We note that in the complex-valued CM criterion, each pair of minima would be replaced by a continuum of minima spanning the full range (0–2π) of allowable system phase.

[7] Well-behaved indicates the absence of common or nearly common subchannel roots.

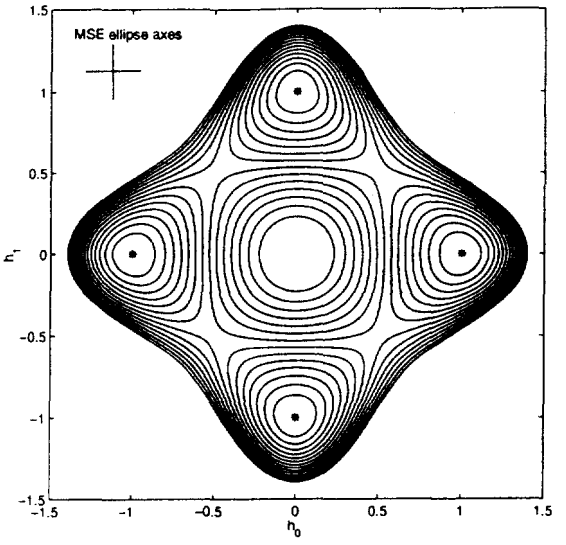


Fig. 10. J_{CM} contours for well-behaved channel $\mathbf{c}_2$ and no noise in combined channel-equalizer ($\mathbf{h}$) space.

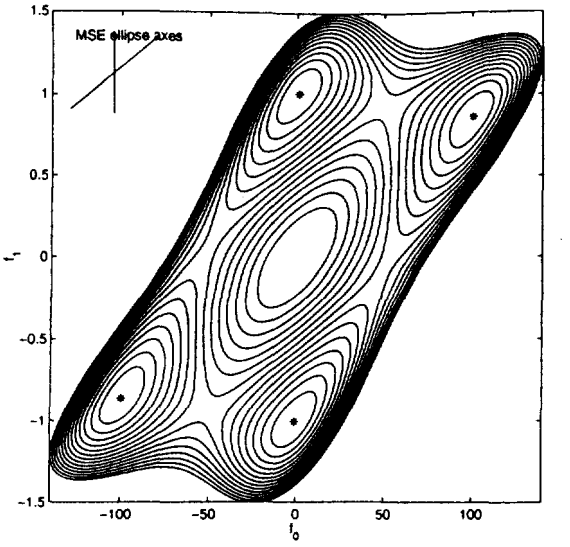


Fig. 12. J_{CM} contours for nearly common subchannel-roots channel $\mathbf{c}_3$ and no noise. Note axis scaling.

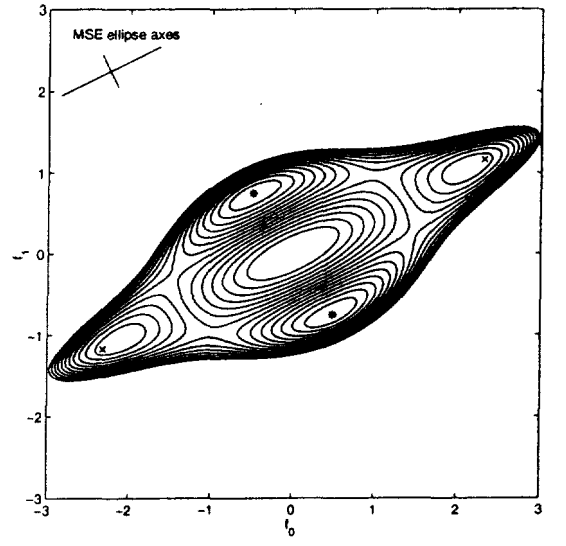


Fig. 11. J_{CM} contours for well-behaved channel $\mathbf{c}_2$ and 20 dB SNR.

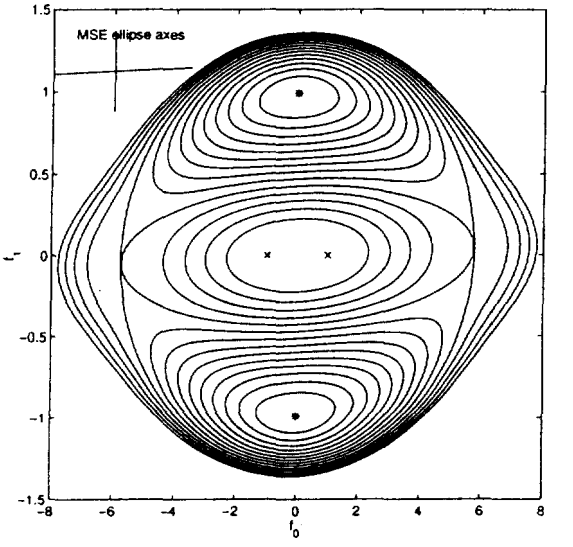


Fig. 13. J_{CM} contours for nearly common subchannel-roots channel $\mathbf{c}_3$ and 20 dB SNR. Note axis scaling.

points on the J_{CM} surfaces in $\mathbf{h}$- and $\mathbf{f}$-spaces when $\mathbf{C}$ is invertible, as it is here in our two-tap example.

3) Additive White Channel Noise: As channel noise is introduced, Fig. 11 indicates that the MSE and CM minima both move toward the origin in $\mathbf{f}$-space. The J_{A} and J_{CM} minima move by different amounts, though, destroying the equivalence that existed between them in the ideal case of Fig. 9. However, the relative proximity of J_{A} and J_{CM} minima evident in Fig. 11 still prompts consideration of J_{CM} as a close proxy for the amalgamated MSE cost J_{A} even when in the presence of channel noise.

4) Common Subchannel Roots: As evidenced by the expression we derived for MSE minima

$$\mathbf{f}^{\dagger} = \left(\mathbf{C}^H\mathbf{C} + \lambda\mathbf{I}\right)^{-1}\mathbf{C}^H\mathbf{h}_{\delta}$$

when $\mathbf{C}^H\mathbf{C}$ has a large condition number, modest values of λ can have significant consequences on $\mathbf{f}^{\dagger}$ (and thus on the J_{A} cost surface). If the two subchannels ($c_0 + c_2 z^{-1}$ and $c_1 + c_3 z^{-1}$) have a nearly common root ($c_3/c_1 \approx c_2/c_0$) then (23) indicates that the column space of $\mathbf{C}$ collapses; thus we expect that one eigenvalue of $\mathbf{C}^H\mathbf{C}$ will be near zero [20]. Figs. 12 and 13 use channel $\mathbf{c}_3$ to demonstrate the cost surface sensitivity to noise in the presence of nearly common subchannel roots. Even under such severe surface deformation, we note that the global J_{CM} minima remain in the vicinity of global J_{A} minima. This further demonstrates the robustness of the relationship between J_{CM} and J_{A}.

5) Channel Undermodeling: In general, under violation of the length condition discussed in Section I-C, no equalizer settings are capable of achieving zero MSE or CM cost. This can be confirmed by extending the length of impulse response $\mathbf{c}_2$ by two samples, thus forming the "undermodeled" channel $\mathbf{c}_4$. (Note that the two extra coef-

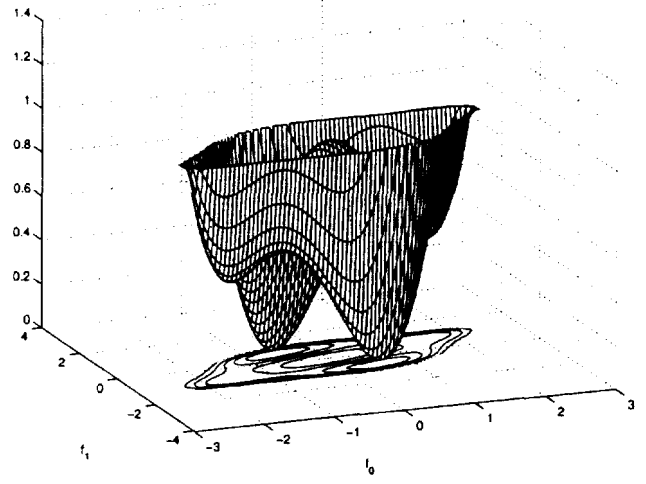


Fig. 14. J_{CM} for undermodeled channel $\mathbf{c}_4$ and no noise.

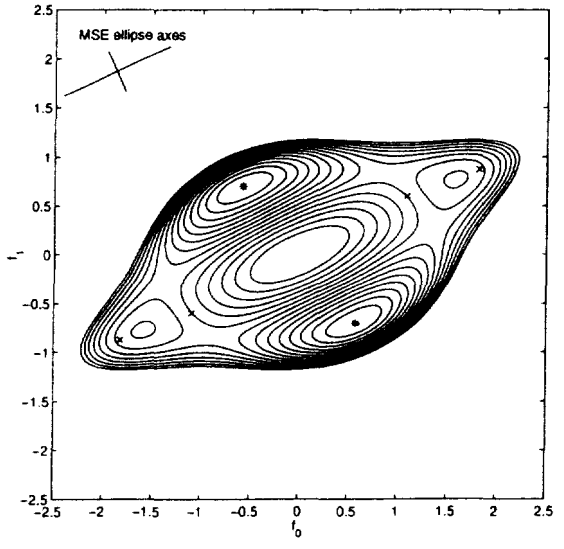


Fig. 15. J_{CM} contours for undermodeled channel $\mathbf{c}_4$ and no noise.

ficients forming $\mathbf{c}_4$ are no larger than any of the coefficients in $\mathbf{c}_2$.) Fig. 14 shows the CM cost surface for this undermodeled channel. Large differences in the heights of local minima demonstrate that the CM cost surface can indeed be significantly multimodal.

Elongating the channel impulse response adds another possibility for the system delay δ and thus increases the number of J_{A} minima (see Fig. 15). Note, however, that the number of CM minima have not changed. More importantly, note that the global CM minima remain close to their MSE counterparts under violations of the length condition.

6) Non-CM Source: The CM source property leading to the ideal zero-cost situation in Figs. 8–10 is violated in constructing the cost surface in Fig. 16. Here, the source is real-valued 32-PAM, which is far from CM. The non-CM property increases the source kurtosis κ_s [defined in (50)] and increases the minimum CM cost relative to that of a CM source. Notice also that the CM cost surface has become "flattened" in the parameter plane. However, as the

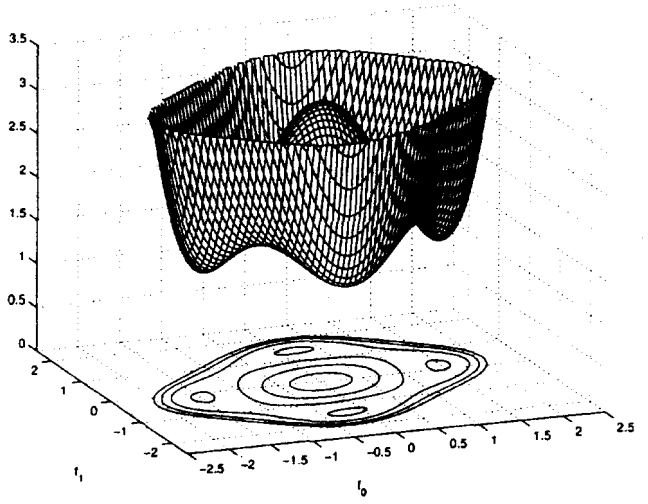


Fig. 16. Effect of source shaping ($\kappa_s = 1.8$) on J_{CM} for channel $\mathbf{c}_1$ in equalizer space with no noise.

CM surface deforms due to a non-CM source, the minima locations remain unchanged.

D. Summary

Our investigations of low-dimensional examples under the following "ideal, zero-cost" conditions:

- no channel noise (i.e., $\lambda = 0$);
- no common subchannel roots (i.e., avoidance of $c_2/c_0 = c_3/c_1$);
- sufficient equalizer length (i.e., $N \geq M - 1$ for $T/2$-spaced FSE's);
- i.i.d., zero-mean, constant-modulus source (circularly symmetric when complex)

showed that, under such conditions, the J_{A} and J_{CM} minima coincide and achieve zero respective cost. Our other examples suggest that modest deviations from the ideal conditions can be tolerated in the following sense: under suitable choice of initialization, a stochastic-gradient minimization of J_{CM} will approximate the performance achieved by the same minimization of J_{A}. We did find, however, that the deformations caused by various violations of the ideal zero-cost conditions are different. In fact, substantial effort has been expended to characterize the performance robustness properties of the CM criterion (as descended by popular gradient descent strategies). Section III catalogs much of this effort.

The previous examples can be used to illustrate and interpret the following observations.

- *Channel noise*: CMA-based blind equalization is typically successful in common noise environments (i.e., $\sigma_s^2 > \sigma_n^2 > 0$). Under modest noise levels, relocation of global minima toward the origin is typically more severe than changes in surface curvature around such minima.
- *Undermodeling of channel length*: Given hardware constraints on equalizer length, residual ISI is unavoidable in practice. Mild contributions from un-

compensated portions of channel response typically result in mild surface deformation.

— *Nearly common subchannel roots*: These seem quite likely as channel length increases (see Fig. 18). Nearly common subchannel roots increase sensitivity to other violations from ideal conditions, but only for sub-optimal CM solutions; global CM minima still exhibit robust performance.

— *Source kurtosis*: Nonuniform (i.e., shaped) symbol distributions often lead to increased source kurtosis. As source kurtosis approaches Gaussian,[8] the surface lifts and flattens. Lifting increases the excess error of stochastic adaptation (e.g., CMA), while flattening reduces its convergence rate. If the source exhibits a Gaussian kurtosis, the minima and saddle points vanish along a rim of the CM surface so that the gradient has solely a radial component. In this case, convergence to desirable settings is practically impossible.

— *Source correlation*: This may occur, e.g., as a result of differential encoding. Small amounts result in slight cost surface deformation. Large amounts cause major problems, such as additional local minima with terrible performance.

— *Non-CM source*: This property is unavoidable in communication systems using multilevel constellations. Though non-CM sources do not alter the minima locations, they raise and flatten the CM surface (as a consequence of increased source kurtosis—see above).

— *Initialization*: The CM surface is unavoidably multimodal. Choice of initialization affects both time-to-convergence and steady-state performance. One approach referred to in the literature suggests initializing the equalizer with a single spike[9] time-aligned with the channel response's center of mass. In this way, crude knowledge of the channel impulse response envelope can be used to aid initialization.

— *Channel time-variation*: We proceed under the global assumption that the channel varies slowly enough in time to be tracked by the CM-minimizing gradient descent algorithm. In the vicinity of a local minimum, the tracking capabilities of any gradient descent scheme can be related to the local curvature.

— *Equalizer tap spacing*: Fractionally spaced equalizers have the ability to perfectly cancel ISI caused by a finite-length channel impulse response. In contrast, a baud-spaced equalizer requires an infinite number of taps for the same capability. Though we admit that this noiseless FIR channel model is rather academic, practical experience offers much evidence for the superiority of fractionally spaced equalization [5].

— *Transient versus steady-state performance*: Dynamic system design is often a tradeoff between transient and steady-state performance. Convergence rate is a transient behavior descriptor; slow convergence is undesired. Excess error (due to a nonvanishing step-size and a nonzero local minimum) is a steady-state feature; abundance of excess error is undesired.

[8] Table 2 presents the values of normalized kurtosis for various sources.

[9] The single-spike initialization has its origins in baud-spaced equalization. Fractionally spaced counterparts are discussed in Section III-B3.

III. CM-Minimizing Equalization

Literature Categorization

Section II presented a tutorial view of the linear equalizer design task and related the minimization of the delay-optimized and phase-indifferent mean-squared recovery error (J_A) to minimization of the CM criterion (J_{CM}). Appendix III presents a bibliography of the literature dealing with the CM criterion and its optimization via steepest gradient descent (such as with CMA). The purpose of this section is to describe our classification scheme in terms of the problem formulation and the examples of the preceding section. We also take this opportunity to cite certain papers as recommended reading on particular topics.

In addition to the birth of the CM criterion in the early 1980's, highlights in its analytical history include:

— the establishment of "perfect" conditions under which a gradient descent of the CM cost surface results in asymptotically perfect symbol recovery, i.e., "global convergence";

— confirmation that, under slightly imperfect conditions, the CM minima remain in the vicinity of the MSE minima for various choices of delay and sign;

— recognition that, due to performance differences between CM minima under less-than-perfect conditions, initialization may be critical to acceptable transient and steady-state behavior.

The "perfect" global-convergence conditions referred to in these statements differ in detail between the baud- and fractionally spaced cases. As discussed in Section I-C, achievement of perfect source recovery devolves into exact solution of a set of simultaneous linear equations when channel noise is absent. Solution of these equations ensures that the transfer function characterizing the baud-spaced system (relating source symbols to equalized soft decisions) achieves that of a pure delay. One requirement on the existence of this perfectly equalizing solution is that the equalizer must have enough degrees of freedom. For a baud-spaced equalizer and an FIR channel, this latter requirement necessitates an equalizer with infinite impulse response (IIR) [31]. For $T/2$-spaced FSE's, on the other hand, an equalizer response length matching (or exceeding) that of the channel proves sufficient [21]. The other requirement for the existence of a perfectly equalizing solution is that the system of equations be well posed. We mean, in an algebraic sense, that the matrix characterizing the linear system of equations must be nonsingular. For baud-spaced equalizers, this nonsingularity condition prohibits nulls in the channel frequency response (which implies, for example, that no FIR channel zeros are tolerated on the unit circle). We henceforth refer to satisfaction of this baud-spaced condition as "invertibility." For $T/2$-spaced FSE's, this nonsingularity translates into a

lack of common subchannel roots (see Appendix I-C) and is commonly referred to as "subchannel disparity."

If conditions on the source (e.g., zero-mean, circularly symmetric, white, and sub-Gaussian) are added onto the perfect equalization requirements described in the last paragraph, a gradient descent of the CM criterion will provide asymptotically perfect source recovery from any baud- or fractionally spaced equalizer initialization. In this case the multiple CM minima all have the same depth, i.e., that of an egg carton. The distinctions in global convergence conditions between the baud- and fractionally spaced cases prompt our separation of these two cases. We note that, while analyses of CM-minimizing baud-spaced equalizers have been published since their introduction in 1980, very few analyses of CM-minimizing fractionally spaced equalizers were published before 1990.

The stringency of the global convergence requirements has prompted theoreticians to examine the impact of their violation. For example, what if the FSE length is less than the total channel response but greater than the "significant" portion of the channel response? How are prominent features of the CM cost surface (e.g., stationary point locations, regions of attraction, and heights of local minima) altered as the source is shaped or correlated and/or channel noise power increases and/or channel disparity is lost? While engineering practice desires answers about simultaneous dissatisfaction of all global convergence conditions, theoretical analysis is more likely to move forward by studying individual (or possibly pairwise) violation of these conditions. Therefore, we are encouraged to adopt a set of literature categorizations concerning studies of robustness to violations in each of the four global-convergence conditions (i.e., absence of channel noise, sufficient length, adequate disparity, and use of a zero-mean, white, circular, sub-Gaussian source process).

In Section II-C we noted that the CM and MSE error surfaces are quite similar in the vicinity of the CM local minima. This relationship implies that the local behaviors of their stochastic gradient descent minimizers (e.g., CMA and LMS, respectively) should be closely related. As a result, we are encouraged to use key behavioral descriptors associated with classical trained-LMS equalization theory as further categories for our literature classification. In particular, we borrow excess MSE (i.e., misadjustment[10]) and convergence rate.

While the CM and MSE criterion are comparable in a local context, their global characteristics are strikingly different. Recall the multimodality of the CM cost surface (e.g., see Figs. 8 and 14). As noted earlier, a good gradient-descent initialization may be necessary to ensure convergence to a "good" local minimum as well as to avoid temporary local capture by saddle points. In contrast, consider the trained-LMS cost surface: a unimodal elliptical hyper-paraboloid. Its unimodality obviates the need for a clever initialization strategy (assuming the training delay has been chosen). In fact, the LMS equalizer is often initialized by zeroing the parameters.[11] If we consider delay-selection as part of the initialization of trained LMS, however, we find many similarities with the equalizer parameter initialization of CMA. Specifically, the choice of training delay bounds asymptotic LMS performance, and, in conjunction with the equalizer initialization, LMS time-to-convergence. Conversely, CMA equalizer initialization determines (asymptotic) system delay. With these thoughts in mind, we add surface topology and initialization strategy as literature categories under the heading of gradient descent behavior.

In summary, the classification scheme we adopt for our literature review uses a total of 11 labels within the three main categories discussed above.

1) Equalizer tap spacing:
   - **B** baud-spaced;
   - **F** fractionally spaced.
2) Global convergence criteria dissatisfaction:
   - **P** perfect; no noise, sufficient length, adequate disparity/invertibility, and zero-mean, white, circular sub-Gaussian source;
   - **N** noise present;
   - **L** equalizer length inadequate;
   - **D** disparity/invertibility lost or threatened;
   - **S** source shaped or correlated.
3) Gradient descent algorithm behavior:
   - **E** excess error (due to nonvanishing step-size);
   - **R** rate of convergence;
   - **T** topology of cost surface;
   - **I** initialization strategy.

The remainder of this section is organized by the categorization above. Each of the 11 labels is discussed using selected citations drawn from Appendix III.

Because the focus of this paper is the CM criterion in a blind linear equalizer application, we have not considered work that

1) principally deals with algorithm modifications (e.g., normalized, least-squares, Newton-based, block, anchored, or signed CMA) that may alter the (effective) cost function surface shape;
2) infers behavior principally from simulation studies with no connection made to the CM cost function;
3) principally addresses applications other than linear equalization (e.g., beamforming, source separation, interference cancellation, channel identification, depolarization, or decision-feedback equalization).

Though some of our citations do involve the categories above, we have chosen to include them because they contain a substantial amount of directly relevant material as well.

[10] Misadjustment is defined as the ratio of excess MSE to minimum MSE.

[11] Initializing CMA at the origin is unwise due to the zero-valued CM-cost gradient there.

We do not provide a synopsis of each citation in the bibliography. Rather, we propose the abstracts of each paper as a source for synopses and provide a postscript bibliography that includes abstracts at http://backhoe.ee.cornell.edu/BERG/bib/CM_bib.ps.

A. Equalizer Tap Spacing

Practically speaking, the equalizer tap spacing refers to the rate at which the received signal is sampled and processed by the equalizer. In creating a discrete linear system model, the tap spacing determines the delay time of the equalizer difference equation. Using T to denote the source symbol interval, baud- or T-spaced FIR equalizers use a unit delay of T seconds in their tapped delay line. Fractionally spaced equalizers use a tap spacing less than T. The most common fractional tap spacing is $T/2$ s. In the bibliography in Section V, approximately two-thirds of the citations cover baud-spaced equalization, while the remaining one-third cover fractionally spaced equalization.

1) Baud-Spaced Equalization: The pioneering paper introducing the CM criterion for a complex-valued source [29] considers baud-spaced equalization only.

Conditions assuring global convergence of a baud-spaced equalizer updated via CMA are: i) no channel noise, ii) infinite impulse response equalizer, iii) no nulls in channel frequency response (i.e., no FIR channel zeros on the unit circle), and iv) a zero-mean, independent (and circularly symmetric if complex-valued) finite-alphabet source with sub-Gaussian kurtosis.

The first proof of global convergence for CMA in adapting a baud-spaced equalizer relied on a doubly infinite equalizer parameterization which allowed any combined channel-equalizer impulse response [31]. This allows convergence study in the combined channel-equalizer space, which has analytical advantages.

2) Fractionally Spaced Equalization: Original motivations for the use of fractional rather than baud spacing included: insensitivity to sampling phase; ability to function as a matched filter; ability to compensate for severe band-edge delay distortion; and reduced noise enhancement [5]. Fractionally spaced equalizers have nearly dominated practice since the 1980's [28]. One feature of fractionally spaced equalizers—virtually unnoticed until the 1990's—was the possibility that under ideal conditions a fractionally spaced equalizer of finite time-span could perfectly equalize an FIR channel [1]. As noted in [21], this suggests the same connection of equalizer parameters to the combined channel-equalizer parameters exploited in [31] and therefore confirms the potential for global convergence of a CM-minimizing fractionally spaced equalizer.

Conditions assuring global convergence of a $T/2$-spaced FSE updated each baud interval via CMA are: i) no channel noise, ii) equalizer time span matching or exceeding that of the FIR channel, iii) no reflected zeros in the $T/2$-sampled FIR channel transfer function, and iv) a zero-mean, independent (and circularly symmetric if complex-valued) finite-alphabet source with sub-Gaussian kurtosis.

These global convergence-inducing conditions do not include restriction to a constant modulus source, which was included among the ideal zero-cost conditions of Section II-D.

The first global convergence proofs for fractionally spaced CMA which do not simply rely on the extension of the baud-spaced arguments in [31] appear in [32].

B. Gradient Descent Algorithm Behavior Theory

The algorithm that performs a stochastic gradient descent of J_{CM} is often referred to as CMA

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \mathbf{r}_n^* y_n \left(\gamma - |y_n|^2\right). \tag{24}$$

Equation (24) is written in terms of the (fractionally sampled) regressor vector at time n

$$\mathbf{r}_n = \left[r_n^{\mathrm{odd}}, \cdots, r_{n-(N-1)}^{\mathrm{odd}}, r_n^{\mathrm{even}}, \cdots, r_{n-(N-1)}^{\mathrm{even}}\right]^t. \tag{25}$$

the equalizer parameter vector $\mathbf{f}_n$ at time index n, the equalizer output y_n, a step-size μ, and the squared source-modulus γ (also referred to as the dispersion constant).

The study of dynamic systems, such as CMA, is often divided into transient and steady-state stages. Convergence rate is the dominant transient performance descriptor in classical LMS theory. MMSE and excess MSE (and their dimensionless ratio, misadjustment = EMSE/MMSE) are the dominant steady-state performance descriptors. Therefore, we consider their CM counterparts here.

Though initialization is not a major concern for the unimodal cost functions of MSE-minimizing equalizers (with preselected delay and phase), it is an unavoidable issue for CM-minimizing equalizers due to the multimodal topology of their associated cost surface. Though initialization strategies exist, none have been proven 100% successful in practice.

1) Convergence Rate: For trained LMS, the convergence rate (or geometric decay factor) of the sum-squared parameter error (and squared recovery error) is approximately bounded above and below by one minus twice the product of the step-size and the smallest and largest eigenvalues, respectively, of the received-signal's autocovariance matrix (i.e., $1 - 2\mu\lambda_{\min} > 1/\tau > 1 - 2\mu\lambda_{\max}$). This arises because the underlying quadratic cost function has the same Hessian, or curvature, across its entire surface. In contrast, the multimodal CM cost function has a Hessian that varies across its surface. Early convergence rate studies addressed this variation in convergence rate across the CM cost surface by focusing on convergence rate descriptors in various regions, such as far from minima and near minima [34].

Referring to Fig. 9, initialization near $[f_0, f_1] = [2.5, 0]$ will lead to a small-stepsize gradient-descent trajectory that passes through the neighborhood of a saddle point. An example displaying multiple temporary saddle-captures appears in [35]. We believe this saddle capture phenomenon to be the source of the folklore that considers CMA to be slowly converging.

A lower bound on the initialization-independent convergence rate is impossible with the multimodal CM surface due to potential of indefinite-term capture by saddle points.

In the neighborhood of a local minimum, the curvature of CMA's cost surface can be directly related to that of trained-LMS [36]. Thus, the LMS convergence rate expression can be used in a traditional manner (e.g., [23]) to provide limits on the channel tracking[12] capabilities of CMA.

2) Excess Cost at Convergence: In realistic situations, it is impossible to zero the update of a nonvanishing-stepsize stochastic gradient descent algorithm, even at the optimum solution. With trained LMS or CMA, this undying perturbation may be a result of channel noise or residual ISI. With CMA, the nonzero update may also be the result of a non-CM source. The effect of a nonvanishing equalizer update is an asymptotic MSE level higher than that attained by the optimum fixed equalizer. This is directly related to the lifting effect that a non-CM source has on the CM cost surface, which is evident in Fig. 16.

In addition to the factors determining the excess MSE of trained LMS (i.e., stepsize, minimum achievable cost, equalizer length, and received signal power) CMA also has a term dependent on the source kurtosis.

Excess MSE of fixed (small) step-size CMA due to a non-CM source is analyzed in [37].

Figs. 8 and 16 show the effect of changing the source from constant to nonconstant modulus while simultaneously satisfying all of the global convergence conditions. Though the CM minima rise in height, they remain in the same locations in the equalizer parameter plane. As long as the source is kept sub-Gaussian, a (pure) gradient descent algorithm would still be able to asymptotically achieve perfect symbol recovery.

3) Initialization: As noted in the examples of Section II-C and illustrated in Figs. 11 and 15, the presence of noise or channel undermodeling causes some CM minima to achieve better performance than others.

Under violation of the conditions ensuring global convergence, choice of initialization determines asymptotic performance.

Two initialization strategies are common in the literature and in practice: spike based or matched filter. The single-spike initialization promoted in [29] for baud-spaced CMA is characterized by one nonzero equalizer tap, usually located somewhere in the central portion of the equalizer tapped-delay line. For $T/2$-spaced CMA, a suitable extension of the single-spike idea might be a "double-spike" initialization, whereby two adjacent taps are initialized nonzero. In the frequency domain, double-spike initialization has a lowpass characteristic, a property also shared by the transmitter's pulse-shaping filter. In a mild-ISI environment, one might even consider initializing the FSE with an impulse response matching the pulse-shaping filter itself, as (in this mild case) this response is close to the expected steady-state equalizer solution (assuming that the FSE is used to accomplish matched filtering at the receiver).

[12] In many practical implementations, such as those with low ambient noise levels, CMA lowers the symbol error rate to a level suitable for decision-directed LMS (DD-LMS) to take over. Due to its lower excess error, DD-LMS is preferred for tracking the slow channel variations. In low-SNR situations, however, such as those that may arise with a coded system, the tracking ability of CMA might prove important due to the potential infeasibility of DD-LMS.

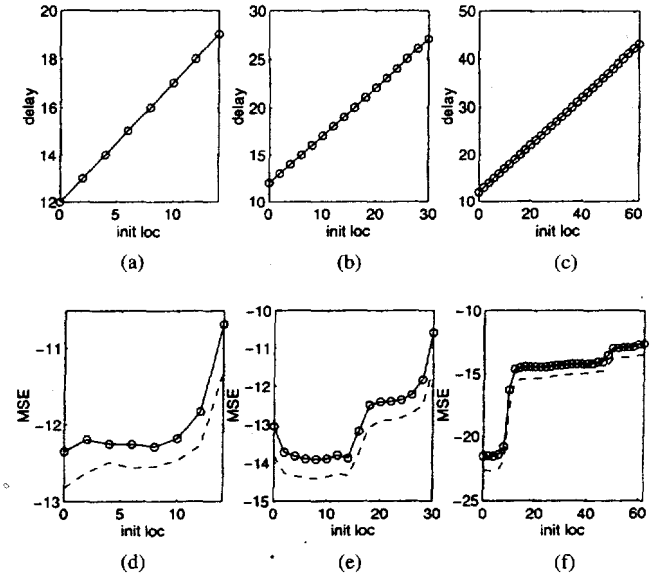


Fig. 17. (a)–(c) CMA's achieved system delay as a function of double-spike location for an equalizer with length 16, 32, 64, respectively. (d)–(f) CMA's asymptotic MSE performance (solid) compared to same-delay MMSE performance (dashed) for an equalizer with length 16, 32, 64, respectively.

All of the initialization techniques above still require a selection of delay, i.e., spike positioning within the equalizer time span. This delay choice is intimately connected to the delay-choice in trained-LMS equalization in the following way: CMA tends to converge to minima with the same group delay as its initialization. Fig. 17 provides evidence for this claim using double-spike initializations of $T/2$-spaced CMA on the SPIB microwave channel shown in Fig. 1 under 50 dB SNR and a QPSK source. Note the (affine) linear correspondence between double-spike position and asymptotically achieved system delay. Another interesting characteristic of Fig. 17, seen after comparing subplots Fig. 17(d)–(f) to Fig. 6, is its suggestion that *the set of system delays reachable by CMA are best in an MMSE sense.* We offer these last two statements as educated conjectures, as no theoretical proofs yet exist to verify them.

The aforementioned relationship between initialization and channel group delays suggests that prior information about the channel may aid in selection of initialization delay choice. Appendix III notes the existence of other more complicated off-line initialization schemes that leverage such notions.

4) Surface Topology: In Figs. 8 and 14, the "molar" shape of the CM cost surface in two-tap real-valued equalizer space is the same one used in Section II-C to aid in an understanding of CMA's transient and asymptotic performance, as well as to motivate the importance of

initialization. Section II-C also described how deformation of this molar shape occurs with violation of the various ideal zero-cost conditions, and it used this surface-centric view to predict the pertinent effects of these violations.

The three-dimensional "molar" shape typical of the real-valued two-tap equalizer CM cost surface offers a compact visualization of virtually all of the major features of CMA behavior theory, applicable even to longer equalizers.

Surface characterization via gradient and Hessian formulas is provided in [38] for baud-spaced equalizers. Reference [39] offers a more developed topological study of the fractionally spaced CM criterion.

C. Violation of Conditions Ensuring Global Convergence

1) Perfect—All Conditions Satisfied: While Sections III-A1 and III-A2 listed conditions ensuring the global convergence of CMA, their violation is unavoidable in practice.

There exists a set of conditions under which an arbitrarily initialized gradient-descent minimization of the CM criterion results in perfect symbol recovery. These "global convergence" inducing conditions, however, are unconditionally violated—if only modestly—in practice.

2) Channel Noise Present: CM-based blind equalization typically remains successful in common noise environments (i.e., $\sigma_s^2 > \sigma_n^2 > 0$). To recall the cost surface deformations due to noise, compare Figs. 9–11.

When the presence of (modest) channel noise is the only violation of the global convergence conditions, the locations of global CM minima shift toward the origin in equalizer parameter space and the minimum achievable CM cost is increased.

This behavior is strikingly similar to the behavior of the MSE criterion in the presence of channel noise. In fact, under modest amounts of noise, the CM minima remain near the MSE minima [33], [40].

At extremely high noise levels (i.e., $\sigma_n^2 > \sigma_s^2$), the two criteria differ in the following manner: the MSE minima continue to move toward the origin, while the CM minima remain within an annulus outside the origin. This behavior is attributed to the so-called "CMA power constraint" [40].

We have also observed the disappearance of local minima under modest-to-high noise levels [41], especially for channels without much disparity (see Fig. 13).

3) Insufficient Equalizer Length: In order to completely cancel the ISI induced by an arbitrary FIR channel, one requires an IIR baud-spaced equalizer or a sufficiently long FIR fractionally spaced equalizer. In the presence of channel noise, the MSE-optimal equalizer makes a compromise between ISI cancellation and noise gain, and the resulting equalizer impulse response is no longer finite-length, even for fractionally spaced equalizers [5].

In the presence of noise, the (baud- and fractionally spaced) MMSE equalizers have an infinite impulse response, implying that the length of an FIR equalizer should be chosen to capture "enough" of the desired response.

Studies on the effect of violations in the equalizer length condition include [42] in a baud-spaced context, and [43], [44] in a fractionally spaced context. The latter provide evidence of CMA robustness to modest channel undermodeling and include approximate bounds on performance.

As hardware advances permit-increased baud-rate, physical channel delay-spreads remain unchanged, and the relative length of the channel impulse response grows proportionally. To combat ISI, there is a corresponding need to increase equalizer length. Therefore, the desire for higher communication rates will always stress the equalization task. This is a primary justification for the continued development and study of truly simple adaptive equalization algorithms like LMS and CMA.

4) Disparity/Invertibility Lost: As discussed earlier, the set of zero-forcing equalizer design equations becomes poorly conditioned in the presence of deep spectral nulls for baud-spaced equalizers, or in the presence of nearly common subchannel roots for fractionally spaced equalizers. Poor conditioning implies an increased parameter sensitivity to noise and other violations of the global convergence conditions. Fortunately, this parameter sensitivity does not imply a performance sensitivity. In other words, global CMA minima remain robust under a loss of disparity. We note that the same is true for the delay-optimal MMSE solutions.

A near-loss of disparity (for FSE's) or invertibility (for BSE's) dramatically increases the sensitivity of suboptimal CM (and MSE) minima to other violations in the global convergence conditions. However, global CM (and MSE) minima remain robust under these conditions.

The behavior of fractionally spaced CM (and MSE) minima under loss of disparity is explained through the following design procedure. For simplicity, let us assume the absence of noise. 1) Factor the common root(s) out of the subchannels in Fig. 5 and form a new system composed of the common root(s) component and what remains of the multichannel component, connected in series. 2) Design the subequalizers so that the remaining multichannel component approximates the inverse of the common root(s) component. At this point, the cascaded system should approximate a pure delay. This procedure closely describes the construction of the MMSE or CM-optimal equalizers under a loss of disparity [33]. We describe this idea more formally in Appendix I-C.

There are a number of reasons that we expect the presence of nearly common subchannel roots, e.g., nearly reflected[13] $T/2$-spaced roots, in realistic situations. Looking at Fig. 18, which portrays the roots of the length 300 $T/2$-sampled SPIB channel whose impulse response appears in Fig. 1 and whose response we consider to be "typical," one notices the apparent plethora of nearly reflected roots. Similarly, one might realize that a long FIR approximation to a pole[14] in the physical channel would also generate nearly reflected roots. These reasons suggest the likelihood

[13] Common subchannel roots have been shown to be identical to $T/2$-spaced channel roots reflected across the origin [25].

[14] A degree-N polynomial forming a close approximation to a single pole can be constructed using N roots on a ring in the complex plane with a radius equal to the pole magnitude. The roots are spaced at $N+1$ equal intervals on the ring with the exception that there exists no root at the location of the approximated pole.

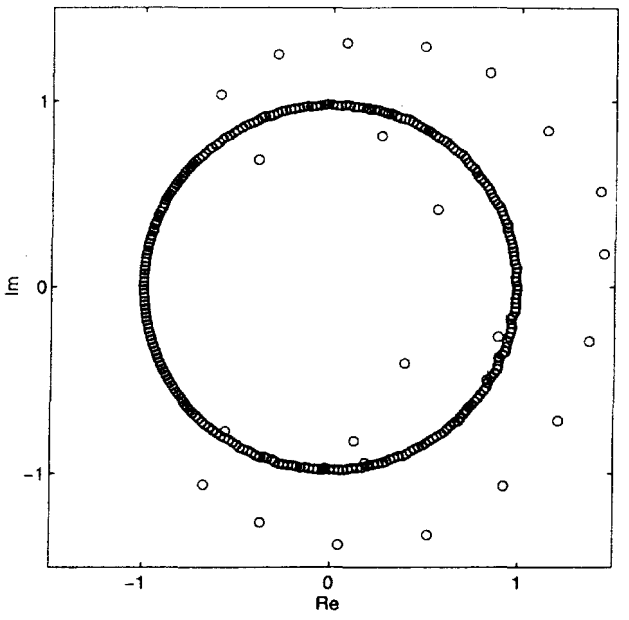


Fig. 18. Roots of $T/2$-sampled SPIB terrestrial microwave channel #3.

of nearly common subchannel roots in realistic situations. See [2] for further discussion on the existence of reflected roots in physical systems and its negative implications on second-order-statistics based blind equalization.

5) Shaped or Correlated Source: Source shaping, encouraged by a potential increase in coding gain (e.g., see [3]), has the effect of making the source symbol distribution more Gaussian. As far as our problem is concerned, it has the practical effect of raising the kurtosis. Increases in source kurtosis, as long as they remain sub-Gaussian, do not affect the locations of CM local minima. However, they are known to flatten the CM cost surface in all but the radial direction, making CMA's convergence to the minima slower (and in the limiting Gaussian case, impossible). In addition, increases in source kurtosis have been shown to raise the CM surface (see Fig. 16), thus increasing the excess asymptotic error levels achieved by nonvanishing-step-size stochastic gradient algorithms.

Recall that non-CM sources also have kurtoses greater than one. To put source shaping in perspective, Table 2 presents the kurtosis of popular source alphabets along with the limiting Gaussian values. Note that a shaped source has the potential for raising the kurtosis far past that of a dense (uniform) constellation like 1024-QAM.

For shaped sources with near-Gaussian kurtoses, the CM cost surface is raised and flattened, therefore unsuited to stochastic gradient descent.

Source correlation results from the use of certain types of coding (e.g., differential encoding) or under particular operational circumstances [45], [46]. Moderate amounts of source correlation may shift the locations of local minima. Large amounts of correlation may cause additional (false) minima to appear in the CM cost surface. Recall that any amount of source correlation violates the CM global convergence requirements. The most thorough studies on the effects of shaped and/or correlated sources appear in [39] and [46].

Table 2 Normalized Kurtoses for Various Source Distributions

real-valued alphabet	kurtosis	complex-valued alphabet	kurtosis
uniform BPSK	1	uniform M-PSK	1
uniform 4-PAM	1.64	uniform 16-QAM	1.32
uniform 8-PAM	1.762	uniform 64-QAM	1.381
uniform 16-PAM	1.791	uniform 256-QAM	1.395
uniform 32-PAM	1.798	uniform 1024-QAM	1.399
Gaussian	3	Gaussian	2

Table 3 Annotations Used in the Bibliography and Their Interpretions

Symbol	Meaning	Comments
P	Perfect Equalization	Addresses the case where any pure delay is achievable; global convergence
N	Noise	Addresses the effects of noise
L	Length	Addresses the effects of equalizer length
D	Disparity	Addresses the effects of channel disparity
S	Source	Addresses the effects of sources that are shaped, non-constant modulus or correlated
F	FSE	Fractionally-spaced equalization context
B	BSE	Baud-spaced equalization context
I	Initialization	Discusses initialization procedures for adaptive implementations
E	Excess error	Discusses sources of excess error in adaptive implementations due to nonvanishing stepsize
R	Convergence Rate	Discusses convergence rate of adaptive implementations
T	Surface Topology	Studies topology of error surface

As a final note, we point out that the global convergence conditions for complex-valued implementations of the CM criterion specify a circularly symmetric source, i.e., $E\{s_n^2\} = 0$. Studies have shown that violations of this requirement (e.g., from the use of a real-valued source with a complex-valued channel and/or equalizer) can result in the appearance of undesired CM minima [47].

D. Additional Information

These descriptions of the literature categorization have prepared the reader to utilize the annotated bibliography in Appendix III, which provides an in-depth look into the CM literature. Each entry in the bibliography is annotated with boldface letters that indicate the classification of its content (see Table 3). A postscript file containing the abstracts of papers in this list is provided available from: http://backhoe.ee.cornell.edu/BERG/bib/CM_bib.ps.

We also recommend The BERGULATOR, a public-domain MATLAB-5 based software environment which allows for experimentation with the CM criterion and various implementations of CMA. It can be used, for example, to generate all contour plots in this paper. The BERGULATOR was written by P. Schniter of Cornell University's Blind Equalization Research Group (CU-BERG) and is available from the following web site: http://backhoe.ee.cornell.edu/BERG/.

APPENDIX I
FRACTIONALLY SPACED SYSTEM MODEL

We denote the combined LTI channel and pulse-shape impulse response by $c(t)$ and the baseband additive channel noise process by $w(t)$. The continuous-time baseband representation of the waveform seen by the receiver can then be described by

$$r(t) = \sum_{n=-\infty}^{\infty} s_n c(t - nT - t_0) + w(t) \tag{26}$$

for symbol sequence $\{s_n\}$, baud interval T, and arbitrary time delay t_0. Sampling[15] the received signal every $T/2$ s at the receiver, we denote the sampled received sequence by

$$r\left(k\frac{T}{2}\right) = \sum_{n=-\infty}^{\infty} s_n c\left(k\frac{T}{2} - nT - t_0\right) + w\left(k\frac{T}{2}\right). \tag{27}$$

The output x_k of a length $2N$ FIR FSE with tap spacing of $T/2$ can be written as a $T/2$-rate convolution with the sampled received sequence

$$x_k = \sum_{i=0}^{2N-1} f_i r\left((k-i)\frac{T}{2}\right). \tag{28}$$

The choice of an even number of equalizer taps is chosen for notational simplicity. Now suppose that only the "odd" fractionally spaced equalizer output samples are retained in a decimation by two (i.e., $k = 2n+1$ for $n = 0, 1, 2, \cdots$). The decimated equalized output sequence y_n^{odd} then becomes

$$y_n^{\text{odd}} = x_{2n+1} \tag{29}$$

$$= \sum_{i=0}^{2N-1} f_i r\left(nT - i\frac{T}{2} + \frac{T}{2}\right) \tag{30}$$

$$= \sum_{i=0}^{N-1} \left(f_{2i} r\left((n-i)T + \frac{T}{2}\right) + f_{2i+1} r((n-i)T)\right). \tag{31}$$

Note that a similar procedure can be carried out for even-indexed output sampling (i.e., $k = 2n$ and $y_n^{\text{even}} = x_{2n}$). An illustration of the setup described above appears in Fig. 3.

A. Multichannel Model

From (31) we observe that the decimated output y_n^{odd} can be considered the sum of two baud-spaced convolutions

$$y_n^{\text{odd}} = \sum_{i=0}^{N-1} \left(f_i^{\text{even}} r_{n-i}^{\text{odd}} + f_i^{\text{odd}} r_{n-i}^{\text{even}}\right) \tag{32}$$

where

$$f_n^{\text{even}} = f_{2n} \qquad f_n^{\text{odd}} = f_{2n+1} \qquad r_n^{\text{even}} = r(nT)$$

and

$$r_n^{\text{odd}} = r\left(nT + \frac{T}{2}\right). \tag{33}$$

We refer to r_n^{even} and r_n^{odd} as the "even" and "odd" received sequences and to f_n^{even} and f_n^{odd} as the "even" and "odd" subequalizers, respectively.

Defining the even and odd baud-rate channel response samples

$$c_n^{\text{even}} = c(nT - t_0) \quad \text{and} \quad c_n^{\text{odd}} = c\left(nT + \frac{T}{2} - t_0\right) \tag{34}$$

and channel noise samples $w_n^{\text{even}} = w(2n(T/2))$ and $w_n^{\text{odd}} = w((2n+1)(T/2))$ (for nonnegative integers n), we can confirm that they are related to the received subsequences in a straightforward manner

$$r_n^{\text{even}} = \sum_l s_l c_{n-l}^{\text{even}} + w_n^{\text{even}} \tag{35}$$

$$r_n^{\text{odd}} = \sum_l s_l c_{n-l}^{\text{odd}} + w_n^{\text{odd}}. \tag{36}$$

These expressions allow us to rewrite the decimated equalizer output in terms of the baud-spaced symbol sequence.

It is important to note that the arbitrary delay t_0 has been incorporated into our definitions of the channel response samples. This implies that the "even" and "odd" subchannel classifications are merely notational and have no real physical significance. Furthermore, the inclusion of arbitrary delay implies that our convention of retaining the odd-indexed (as opposed to the even-indexed) decimated equalizer output samples also lacks practical significance. In this spirit, we drop the "odd" notation on y_n^{odd} and simply refer to the baud-spaced system output samples as y_n. Here we are seeing evidence for the inherent baud-synchronization capabilities of an FSE (not characteristic of BSE's).

Substituting the received subsequence expressions (35) and (36) into (32)

$$y_n = \sum_{i=0}^{N-1} f_i^{\text{even}} \left(\sum_l s_l c_{n-i-l}^{\text{odd}} + w_{n-i}^{\text{odd}}\right) + \sum_{i=0}^{N-1} f_i^{\text{odd}} \left(\sum_l s_l c_{n-i-l}^{\text{even}} + w_{n-i}^{\text{even}}\right) \tag{37}$$

$$= s_n \star (f_n^{\text{even}} \star c_n^{\text{odd}} + f_n^{\text{odd}} \star c_n^{\text{even}}) + f_n^{\text{even}} \star w_n^{\text{odd}} + f_n^{\text{odd}} \star w_n^{\text{even}} \tag{38}$$

where the "$\star$" indicates convolution. The relationships between the source, noise, subequalizers, and subchannels described above appears in the multichannel model of Fig. 5.

Consider for a moment the noiseless case. The impulse response h_n from transmitted source to baud-spaced equalizer output follows immediately from consideration of s_n as the Kronecker delta sequence δ_n. Thus, we conclude that

$$h_n = f_n^{\text{even}} \star c_n^{\text{odd}} + f_n^{\text{odd}} \star c_n^{\text{even}}. \tag{39}$$

[15] The noise and channel are considered band-limited assuming antialias filtering is done prior to $T/2$-spaced sampling at the receiver.

This impulse response leads directly to a transfer function $H(z^{-1})$ with unit delay (z^{-1}) of duration T

$$H(z^{-1}) = F_{\text{even}}(z^{-1})C_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})C_{\text{even}}(z^{-1}). \quad (40)$$

Note that the perfect zero-forcing system $H(z) = z^{-\delta}$ (with nonnegative integer delay δ), leads to the Bezout relationship [8]

$$z^{-\delta} = F_{\text{even}}(z^{-1})C_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})C_{\text{even}}(z^{-1}). \quad (41)$$

B. Multirate Model

To show that the multirate model of Fig. 4 also originates from the fractionally spaced communication system of Fig. 3, we show that the fractionally spaced equalizer output $\{x_k\}$ in (28) can be written in terms of a zero-filled version of the source sequence $\{a_k\}$

$$a_k = \begin{cases} s_{k/2}, & \text{for } k \text{ even} \\ 0, & \text{for } k \text{ odd} \end{cases} \quad (42)$$

as depicted in Fig. 4. Rewriting (27) as

$$r\left(k\frac{T}{2}\right) = \sum_{l=-\infty}^{\infty} a_l c\left((k-l)\frac{T}{2} - t_0\right) + w_k \quad (43)$$

we see upon its substitution into (28) that

$$x_k = \sum_{i=0}^{2N-1} f_i\left(\sum_l a_l c\left((k-i-l)\frac{T}{2} - t_0\right) + w_{k-i}\right) \quad (44)$$

$$= \sum_{i=0}^{2N-1} f_i\left(\sum_l a_l c_{k-i-l} + w_{k-i}\right) \quad (45)$$

$$= f_k \star (a_k \star c_k + w_k) \quad (46)$$

where the fractionally spaced channel response samples c_k are defined such that $c_k = c(k(T/2) - t_0)$.

At this point we can observe that, in the noiseless case, the fractionally spaced system impulse response h_k^{FS} becomes

$$h_k^{\text{FS}} = f_k \star c_k. \quad (47)$$

Note from (39) that only half of the terms in the fractionally spaced impulse response (47) are directly relevant to the system output since the fractionally spaced output $\{x_k\}$ is later decimated by two.

C. The Subchannel Disparity Condition

The Bezout equation (41) leads directly to the perfect equalization requirement concerning subchannel roots. Specifically, for the existence of a (finite-length) zero-forcing equalizer, the subchannel polynomials $C_{\text{even}}(z^{-1})$ and $C_{\text{odd}}(z^{-1})$ must not share a common root.

The existence of perfectly equalizing sub-equalizer polynomials $F_{\text{even}}(z^{-1})$ and $F_{\text{odd}}(z^{-1})$ implies that (41) can be satisfied. For example, if the subchannels share one root, a common polynomial $G(z^{-1}) = g_0 + g_1 z^{-1}$ can be factored out of both $C_{\text{even}}(z^{-1})$ and $C_{\text{odd}}(z^{-1})$, leaving $\overline{C}_{\text{even}}(z^{-1})$ and $\overline{C}_{\text{odd}}(z^{-1})$, respectively. The perfect equalization relationship would then become

$$z^{-\delta} = G(z^{-1})\left(F_{\text{even}}(z^{-1})\overline{C}_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})\overline{C}_{\text{even}}(z^{-1})\right) \quad (48)$$

but this is contradicted by the fact that there is no finite-length polynomial $F_{\text{even}}(z^{-1})\overline{C}_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})\overline{C}_{\text{even}}(z^{-1})$ that when multiplied by $G(z^{-1})$ results in the delay operator $z^{-\delta}$.

However, $F_{\text{even}}(z^{-1})$ and $F_{\text{odd}}(z^{-1})$ can be chosen so that (48) is approximated, in which case the following relationship is satisfied:

$$z^{-\delta}G^{-1}(z^{-1}) \approx F_{\text{even}}(z^{-1})\overline{C}_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})\overline{C}_{\text{even}}(z^{-1}). \quad (49)$$

In other words, the FSE combines with the noncommon-root component of the channel to approximate the (IIR) inverse of the (T-spaced) common root component.

D. On The Independence of Fractionally Sampled Channel Noise

A typical assumption on the (baseband equivalent) channel noise $w(t)$ is that it is well modeled by a zero-mean, circularly symmetric Gaussian process [10]. In many situations $w(t)$ is also assumed to have a flat wideband power spectrum. Does this imply that the fractionally sampled noise process $\{w_k\}$ will also be white? Under these conditions, $\{w_k\}$ will only be white when the anti-alias filters prior to $T/2$-spaced sampling satisfy a rate $2/T$ Nyquist criterion. In practice, this criterion is satisfied by anti-alias filters that are power-symmetric about the frequency $1/T$ Hz. If, for example, the filtering prior to equalization is matched to the pulse shape of the transmitted signal, then fractionally sampled $\{w_k\}$ will not be white.

Appendix II
The CM Cost Function

Below we provide the general formulation of the CM cost function for a complex i.i.d. zero-mean source and complex baseband channel in additive white zero-mean noise. We will assume that each member of the symbol alphabet is equiprobable in the source sequence. Furthermore, we also assume that the receiver sampling clock is frequency synchronous (a fixed time offset is allowed) with the source symbol clock. In practice, this is a reasonable assumption since the symbol clock can often be extracted by computing the square magnitude of the received signal (commonly known as envelope detection). Given these assumptions, we follow the general formulation of the CM cost function with expressions for the specific cases of PAM, PSK, and QAM input signals.

In addition to the previously introduced notation we will use the following definitions:

$$\kappa_s = \frac{E\{|s_n|^4\}}{\sigma_s^4}, \text{ the normalized kurtosis of } \{s_n\} \quad (50)$$

$$\gamma = \frac{E\{|s_n|^4\}}{\sigma_s^2}, \text{ the dispersion constant of } \{s_n\} \quad (51)$$

$$\|\mathbf{h}\|_2^2 = \sum_{n=0}^{P-1} |h_n|^2, \text{ the squared } \ell_2\text{-norm of } \mathbf{h}. \quad (52)$$

Note that $\gamma = \sigma_s^2 \kappa_s$, where $\sigma_s^2 = E\{|s_n|^2\}$. Following the presentation of the FS system model in Section I-C and Appendix I, we can redefine the equalizer output using (4) and (13). This results in

$$y_n = \mathbf{h}^t \mathbf{s}(n) + \mathbf{f}^t \mathbf{w}(n). \quad (53)$$

The CM cost function is

$$J_{\text{CM}} = E\left\{\left(\gamma - |y_n|^2\right)^2\right\} \quad (54)$$

$$= E\{|y_n|^4\} - 2\gamma E\{|y_n|^2\} + \gamma^2$$

$$= E\{|y_n|^4\} - 2\sigma_s^2 \kappa_s E\{|y_n|^2\} + \sigma_s^4 \kappa_s^2. \quad (55)$$

In order to analyze J_{CM}, we will first expand $|y_n|^2$, using (53). For convenience we will temporarily let $A_n = \mathbf{h}^t \mathbf{s}(n)$ and $B_n = \mathbf{f}^t \mathbf{w}(n)$, where $y_n = A_n + B_n$. Using the assumptions of mutually independent zero-mean noise and source sequences, we note that A_n and B_n are also independent and zero-mean, i.e.,

$$E\{A_n\} = \mathbf{h}^t E\{\mathbf{s}(n)\} = 0$$

$$E\{B_n\} = \mathbf{f}^t E\{\mathbf{w}(n)\} = 0$$

and

$$E\{A_n B_n\} = E\{A_n\} E\{B_n\}.$$

With these assumptions, we arrive at

$$E\{|y_n|^2\} = E\{|A_n|^2\} + E\{A_n\} E\{B_n^*\} + E\{A_n^*\} E\{B_n\} + E\{|B_n|^2\}$$

$$= E\{|A_n|^2\} + E\{|B_n|^2\}. \quad (56)$$

Expanding $|A_n|^2$ and $|B_n|^2$, we have that

$$E\{|y_n|^2\} = \sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2. \quad (57)$$

The same approach can be used to examine $E\{|y_n|^4\}$, which leads to the following equation:

$$E\{|y_n|^4\} = E\{|A_n|^4\} + E\{A_n^2\} E\left\{(B_n^*)^2\right\} + 4E\{|A_n|^2\} E\{|B_n|^2\} \quad (58)$$

$$+ E\{B_n^2\} E\left\{(A_n^*)^2\right\} + E\{|B_n|^4\}. \quad (59)$$

Due to space limitations, we omit the details of the derivation of $E\{|y_n|^4\}$ but mention the following properties used in the derivation.

— The second-order terms are relatively easy to compute; they involve summations of source (and noise) terms of the form $E\{s_{n-i} s_{n-l}\}$, $E\{s_{n-i} s_{n-l}^*\}$, or $E\{s_{n-i}^* s_{n-l}^*\}$.

— The fourth-order terms are more difficult to compute, but each of the source (and noise) terms are of the form $E\{s_{n-i} s_{n-l}^* s_{n-m} s_{n-j}^*\}$.

Any of the expectations not involving an even power (two or four) will vanish because the source and noise are both zero-mean and white. After a considerable amount of algebra we arrive at the following expression for $E\{|y_n|^4\}$. Noting that $E\{s_n^2\}$ and $E\{w_n^2\}$ are independent of n, we will denote expectations of this form by $E\{s^2\}$ and $E\{w^2\}$, respectively

$$E\{|y_n|^4\} = \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0, m\neq i}^{P-1} |h_i|^2 \cdot |h_m|^2 + |E\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0, j\neq i}^{P-1} h_i^2 (h_j^*)^2$$

$$+ \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0, m\neq i}^{2N-1} |f_i|^2 |f_m|^2 + |E\{w^2\}|^2 \sum_{i=0}^{2N-1} \sum_{j=0, j\neq i}^{2N-1} f_i^2 (f_j^*)^2$$

$$+ \left(E\{s^2\} \sum_{i=0}^{P-1} h_i^2\right)\left(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2\right)^* + 4\sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 + \left(E\{s^2\} \sum_{i=0}^{P-1} h_i^2\right)^* \cdot \left(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2\right). \quad (60)$$

We define the noise kurtosis κ_w analogous to the source kurtosis κ_s in (50). Substituting (57) and (60) into (55) we have the final expansion of the cost function.

$$J_{\text{CM}} = \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0, m\neq i}^{P-1} |h_i|^2 \cdot |h_m|^2 + |E\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0, j\neq i}^{P-1} h_i^2 (h_j^*)^2$$

$$+ \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0, m\neq i}^{2N-1} |f_i|^2 |f_m|^2 + |E\{w^2\}|^2 \sum_{i=0}^{2N-1} \sum_{j=0, j\neq i}^{2N-1} f_i^2 (f_j^*)^2$$

$$+ \left(E\{s^2\} \sum_{i=0}^{P-1} h_i^2\right)\left(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2\right)^* + 4\sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 + \left(E\{s^2\} \sum_{i=0}^{P-1} h_i^2\right)^* \cdot \left(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2\right) - 2\sigma_s^2 \kappa_s \cdot \left(\sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2\right) + \sigma_s^4 \kappa_s^2. \quad (61)$$

We will now consider how various restrictions on the source and noise simplify this equation.

A. PAM Source, Real-Valued Channel

For PAM the source symbols s_n are real valued so that $E\{|s_n|^2\} = E\{s_n^2\} = \sigma_s^2$. Furthermore, if w_n, $\mathbf{f}$, and $\mathbf{h}$ are real valued, we have $E\{|w_n|^2\} = E\{w_n^2\} = \sigma_w^2$, $f_i^2 = (f_i^*)^2 = |f_i|^2$, and $h_i^2 = (h_i^*)^2 = |h_i|^2$. Thus, we have that, for a real-valued source and real-valued channel, (61) reduces to

$$\begin{aligned} J_{\mathrm{CM}|\mathrm{PAM}} = \kappa_s\sigma_s^4 \sum_{i=0}^{P-1} h_i^4 + 3\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0,\, m\neq i}^{P-1} h_i^2 h_m^2 \\ + \kappa_w\sigma_w^4 \sum_{i=0}^{2N-1} f_i^4 + 3\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0,\, m\neq i}^{2N-1} \\ \cdot f_i^2 f_m^2 + \sigma_s^2\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 + 4\sigma_s^2\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 \\ + \sigma_s^2\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 \\ - 2\sigma_s^2\kappa_s\left(\sigma_s^2\|\mathbf{h}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2\right) + \sigma_s^4\kappa_s^2. \end{aligned}$$

Noting that

$$\sum_{i=0}^{P-1} \sum_{m=0,\, m\neq i}^{P-1} h_i^2 h_m^2 = \|\mathbf{h}\|_2^4 - \sum_{i=0}^{P-1} h_i^4$$

and summing like terms we arrive at

$$\begin{aligned} J_{\mathrm{CM}|\mathrm{PAM}} = \sigma_s^4(\kappa_s - 3) \sum_{i=0}^{P-1} h_i^4 + 3\sigma_s^4\|\mathbf{h}\|_2^4 \\ + \sigma_w^4(\kappa_w - 3) \sum_{i=0}^{2N-1} f_i^4 + 3\sigma_w^4\|\mathbf{f}\|_2^4 \\ + 6\sigma_s^2\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 \\ - 2\sigma_s^2\kappa_s(\sigma_s^2\|\mathbf{h}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2) + \sigma_s^4\kappa_s^2. \end{aligned} \quad (62)$$

Note that if the noise is Gaussian, $\kappa_w = 3$ and the third term in (62) is zero.

1) BPSK Source, Real-Valued Channel: Here we consider the subcase of a BPSK source in a real-valued channel results in further simplifications. For BPSK, $\kappa_s = \sigma_s^4 = 1$, which implies that (62) reduces to

$$\begin{aligned} J_{\mathrm{CM}|\mathrm{BPSK}} = -2 \sum_{i=0}^{P-1} h_i^4 + 3\|\mathbf{h}\|_2^4 + \sigma_w^4(\kappa_w - 3) \\ \cdot \sum_{i=0}^{2N-1} f_i^4 + 3\sigma_w^4\|\mathbf{f}\|_2^4 + 6\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 \\ - 2\left(\|\mathbf{h}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2\right) + 1. \end{aligned} \quad (63)$$

In the absence of noise, (63) is the equation given in [38].

2) Complex-Valued Rotationally Invariant Noise: If we make the assumption that the (complex) noise is rotationally invariant, i.e., $p(w = \rho e^{j\theta}) = p(|w| = \rho)/2\pi$ for all $\theta \in [0, 2\pi]$, then we have that $E\{w^n\} = E\{\rho^n\}E\{e^{jn\theta}\} = 0$, for $n = 1, 2, \cdots$. Using this assumption the cost function reduces to

$$\begin{aligned} J_{\mathrm{CM}|\mathrm{r.i.\ noise}} = \kappa_s\sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0,\, m\neq i}^{P-1} \\ \cdot |h_i|^2|h_m|^2 + |E\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0,\, j\neq i}^{P-1} \\ \cdot h_i^2(h_j^*)^2 + \kappa_w\sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \\ \cdot \sum_{m=0,\, m\neq i}^{2N-1} |f_i|^2|f_m|^2 + 4\sigma_s^2\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 \\ - 2\sigma_s^2\kappa_s(\sigma_s^2\|\mathbf{h}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2) + \sigma_s^4\kappa_s^2 \end{aligned} \quad (64)$$

For the remaining derivations we will make the assumption of rotationally invariant noise.

3) PSK Source: For PSK symbols, $s_n \in \{e^{j2\pi m/2^M}\}$ and $m \in \{0, 1, \cdots, 2^M - 1\}$ (where $j = \sqrt{-1}$), we note that $\sigma_s^2 = E\{|s|^4\} = \kappa_s = \sigma_s^4 = 1$. Thus, (64) simplifies to

$$\begin{aligned} J_{\mathrm{CM}|\mathrm{PSK}} = \sum_{i=0}^{P-1} |h_i|^4 + 2 \sum_{i=0}^{P-1} \sum_{m=0,\, m\neq i}^{P-1} |h_i|^2|h_m|^2 \\ + |E\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0,\, j\neq i}^{P-1} h_i^2(h_j^*)^2 + \kappa_w\sigma_w^4 \\ \cdot \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0,\, m\neq i}^{2N-1} |f_i|^2|f_m|^2 \\ + 4\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 - 2(\|\mathbf{h}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2) + 1. \end{aligned} \quad (65)$$

4) QAM Source: For 90° rotationally invariant QAM (i.e., for every member q_m in the QAM alphabet, $\{jq_m, -q_m, -jq_m\}$ are equally likely members of the alphabet), we have that $E\{s^2\} = 0$ and (64) reduces to

$$\begin{aligned} J_{\mathrm{CM}|\mathrm{QAM}} = \kappa_s\sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0,\, m\neq i}^{P-1} |h_i|^2 \\ \cdot |h_m|^2 + \kappa_w\sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \\ \cdot \sum_{m=0,\, m\neq i}^{2N-1} |f_i|^2|f_m|^2 + 4\sigma_s^2\sigma_w^2\|\mathbf{h}\|_2^2\|\mathbf{f}\|_2^2 \\ - 2\sigma_s^2\kappa_s(\sigma_s^2\|\mathbf{h}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2) + \sigma_s^4\kappa_s^2. \end{aligned} \quad (66)$$

APPENDIX III
CM-MINIMIZING EQUALIZATION LITERATURE

1) R. A. Axford, Jr., L. B. Milstein, and J. R. Zeidler, "On the misconvergence of CMA blind equalizers in the reception of PN sequences," in *Proc. IEEE Military Commun. Conf.*, Fort Monmouth, NJ, Oct. 2–5, 1994, pp. 281–286. **Categories B, S, R.**

2) ——, "Refined techniques for blind equalization of phase-shift keyed (PSK) and quadrature amplitude modulated QAM

digital communications signals," Ph.D. dissertation, Univ. of California, San Diego, 1995. **Categories B, S, E.**

3) R. A. Axford, Jr., L. B. Milstein, and J. R. Zeidler, "The effects on PN sequences on the misconvergence of the constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 46, pp. 519–523, Feb. 1998. **Categories B, S, R.**
4) N. J. Bershad and S. Roy, "Performance of the 2-2 constant modulus (CM) adaptive algorithm for Rayleigh fading sinusoids in Gaussian noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 1675–1678. **Categories B, N.**
5) C. K. Chan and J. J. Shynk, "Stationary points of the constant modulus algorithm for real Gaussian signals," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 38, pp. 2176–2181, Dec. 1990. **Categories B, S, T.**
6) W. Chung and J. P. LeBlanc, "The local minima of fractionally spaced CMA blind equalizer cost function in the presence of channel noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Seattle, WA, May 1998, pp. 3345–3348. **Categories F, T, N.**
7) R. Cusani and A. Laurenti, "Evaluation of the constant modulus algorithm in blind equalization of three ray multipath fading channels," *European Trans. Telecommun. Related Tech.*, vol. 6, no. 2, pp. 187–90, Feb.–Apr. 1995.
8) ——, "Convergence analysis of the CMA blind equalizer," *IEEE Trans. Commun.*, vol. 43, pp. 1304–1307, Feb.–Apr. 1995. **Category B.**
9) Z. Ding, C. R. Johnson, Jr., R. A. Kennedy, and B. D. O. Anderson, "On the ill-convergence of Godard blind equalizers in data communication systems," in *Proc. Conf. Inform. Science and Systems*, Baltimore, MD, Mar. 1989, pp. 538–543. **Categories B, L, T, I.**
10) Z. Ding, C. R. Johnson, Jr. and, R. A. Kennedy, "Admissibility in blind adaptive equalizers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 1707–1710. **Category B.**
11) Z. Ding, "Application aspects of blind adaptive equalizers in QAM data communications," Ph.D. dissertation, Cornell University, Ithaca, NY, 1990. **Categories B, L, T, I.**
12) Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "Local convergence of 'globally convergent' blind adaptive equalization algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Toronto, Ont., Canada, May 14–17, 1991, pp. 1533–1536. **Categories B, L, T.**
13) Z. Ding and C. R. Johnson, Jr., "Existing gap between theory and application of blind equalization," in *Proc. Int. Soc. Optical Engineers*, San Diego, CA, July 1991, pp. 154–165. **Categories B, S, T.**
14) Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, Jr., "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Commun.*, vol. 39, pp. 1313–1327, Sept. 1991. **Categories B, L, T.**
15) Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizers," *IEEE Trans. Signal Processing*, vol. 40, pp. 2425–2432, Oct. 1992. **Categories B, L, T.**
16) Z. Ding, and R. A. Kennedy, "On the whereabouts of local minima for blind adaptive equalizers," *IEEE Trans. Circuits Syst. II*, vol. 39, pp. 119–123, Feb. 1992. **Categories B, L, T.**
17) Z. Ding and C. R. Johnson, Jr., "On the nonvanishing stability of undesirable equilibria for FIR Godard blind equalizers," *IEEE Trans. Signal Processing*, vol. 41, pp. 1940–1944, May 1993. **Categories B, L, T.**
18) Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "Global convergence issues with linear blind adaptive equalizers," in *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994, pp. 60–120. **Categories B, L, T.**
19) Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 45, pp. 650–657, Mar. 1997. **Categories F, P, T.**
20) T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr., and M. Green, "On the robustness of the fractionally spaced constant modulus criterion to channel order undermodeling: Part I," in *Proc. IEEE Signal Processing Workshop Signal Processing Advances Wireless Communications*, Paris, France, Apr. 16–18, 1997, pp. 37–40. **Categories F, L, T.**
21) ——, "On the robustness of the fractionally spaced constant modulus criterion to channel order undermodeling: Part II," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Munich, Germany, Apr. 20–24, 1997, pp. 3605–3608. **Categories F, L, T, E.**
22) T. J. Endres, "Equalizing with fractionally spaced constant modulus and second-order-statistics blind receivers," Ph.D. dissertation, Cornell University, Ithaca, NY, 1997. **Categories F, L, T, E.**
23) S. Evans and L. Tong, "Adaptive channel surfing reinitialization of the constant modulus algorithm," in *Proc. Asimolar Conf. Signals, Systems, and Computers.*, Pacific Grove, CA, Nov. 1997. **Categories B, I.**
24) I. Fijalkow, F. Lopez de Victoria, and C. R. Johnson, Jr., "Adaptive fractionally spaced blind equalization," in *Signal Processing Workshop*, Yosemite National Park, CA, Oct. 2–5, 1994, pp. 257–260. **Categories F, P, T.**
25) I. Fijalkow, J. R. Treichler, and C. R. Johnson, Jr., "Fractionally spaced blind equalization: Loss of channel disparity," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, May 9–12, 1995, pp. 1988–1991. **Categories F, D.**
26) I. Fijalkow, A. Touzni, and C. R. Johnson, Jr., "Spatio-temporal equalizability under channel noise and loss of disparity," in *Proc. Colloque GRETSI Traitment du Signal et des Images*, Sophia Antipolis, France, Sept. 1995, pp. 293–296. **Categories F, N, D.**
27) I. Fijalkow, A. Touzni, and J. R. Treichler, "Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 45, pp. 56–66, Jan. 1997. **Categories F, N, D, T.**
28) I. Fijalkow, C. Manlove, and C. R. Johnson, Jr., "Adaptive fractionally spaced blind CMA equalization: Excess MSE," *IEEE Trans. Signal Processing*, vol. 46, pp. 227–231, Jan. 1998. **Categories F, P, E.**
29) G. J. Foschini, "Equalizing without altering or detecting data (digital radio systems)," *AT&T Tech. J.*, vol. 64, no. 8, pp. 1885–1911, Oct. 1985. **Categories B, P, T, R.**
30) M. R. Frater, R. R. Bitmead, and C. R. Johnson, Jr., "Escape from stable equilibria in blind adaptive equalizers," in *Proc. IEEE Conf. Decision Control*, Tucson, AZ, Dec. 1992, pp. 1756–1761. **Categories B, T, R.**
31) M. R. Frater and C. R. Johnson, Jr., "Local minima escape transients of CMA," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Adelaide, SA, Australia, Apr. 19–22, 1994, pp. 37–40. **Categories B, T, R.**
32) M. R. Frater, R. R. Bitmead, and C. R. Johnson, Jr., "Local minima escape transients by stochastic gradient descent algorithms in blind adaptive equalizers," *Automatica*, vol. 31, no. 4, pp. 637–641, Apr. 1995. **Categories B, T, R.**
33) D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980. **Categories B, T, I, P.**
34) Z. Gu and W. A. Sethares, "A geometrical view of blind equalization," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, Apr. 27–30, 1993, vol. 3, pp. 551–554. **Categories B, T.**
35) K. Hilal and P. Duhamel, "A general form for recursive adaptive algorithms leading to an exact recursive CMA," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, San Francisco, CA, Mar. 1992, pp. 17–20. **Categories B, R.**
36) ——, "A convergence study of the constant modulus algorithm leading to a normalized-CMA and a block-normalized-CMA," in *Proc. European Signal Processing Conf.*, Brussels, Belgium, Aug. 24–27, 1992, vol. 1, pp. 135–138. **Categories B, R.**
37) ——, "Stability and convergence analysis of the constant modulus algorithm. Comments on finite equalization schemes and the stability of the normalized CMA," in *Proc. European Signal Processing Conf.*, Edinburgh, UK, Sept. 13–16, 1994, pp. 724–727.
38) H. Jamali and S. L. Wood, "Error surface analysis for the complex constant modulus adaptive algorithm," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1990, pp. 248–252. **Categories B, T.**
39) H. Jamali, S. L. Wood, and R. Cristi, "Experimental validation of the Kronecker product Godard blind adaptive algorithms," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 26–28, 1992, vol. 1, pp. 1–5. **Categories B, T, N, L.**

40) C. R. Johnson, Jr., S. Dasgupta, and W. A. Sethares, "Averaging analysis of local stability of a real constant modulus algorithm adaptive filter," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 36, pp. 900–910, June 1988. **Category B.**

41) C. R. Johnson, Jr., "Admissibility in blind adaptive channel equalization," *IEEE Contr. Syst. Mag*, vol. 11, pp. 3–15, Jan. 1991. **Categories B, T, I.**

42) C. R. Johnson, Jr., P. C. E. Bennet, J. P. LeBlanc, and V. Krishnamurthy, "Blind adaptive equalizer average stationary point stability analysis with admissibility consequences," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 1992, pp. 727–732. **Categories B, S, L, I.**

43) C. R. Johnson, Jr. and J. P. LeBlanc, "Towards operational guidelines for memoryless-error-function-style blind equalizers," in *Proc. COST 229 Workshop Adaptive Algorithms Commun.*, Bordeaux-Technopolis, France, Sept. 1992, pp. 5–17. **Categories B, S.**

44) C. R. Johnson, Jr., J. P. LeBlanc, and V. Krishnamurthy, "Godard blind equalizer misbehavior with correlated sources," *J. Marocain d'Automat., d'Inform., Signal.*, vol. 2, pp. 1–39, June 1993. **Categories B, S.**

45) C. R. Johnson, Jr. and B. D. O. Anderson, "Godard blind equalizer error surface charateristics: White, zero-mean, binary case," *Int. J. Adpative Contr. Signal Processing*, vol. 9, pp. 301–324, July–Aug. 1995. **Categories B, T, L.**

46) O. W. Kwon, C. K. Un, and J. C. Lee, "Performance of constant modulus adaptive digital filters for interference cancellation," *Signal Processing*, vol. 26, pp. 185–196, Feb. 1992. **Categories B, R, N.**

47) S. Lambotharan, J. Chambers, and C. R. Johnson, Jr., "Attraction of saddles and slow convergence in CMA adaptation," *Signal Processing*, vol. 59, pp. 335–340, June 1997. **Categories F, R, T, P.**

48) M. G. Larimore and J. R. Treichler, "Convergence behavior of the constant modulus algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Boston, MA, Apr. 14–16, 1983, pp. 13–16. **Categories B, R.**

49) ——, "The capture properties of CMA-based interference cancellers," in *Proc. Asilomar Conf. Signals, Systems., and Computers*, Pacific Grove, CA, Nov. 5–7, 1984, pp. 49–52. **Categories B, S, I, N.**

50) ——, "Noise capture properties of constant modulus algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Tampa, FL, Mar. 26–29, 1985, pp. 1165–1168. **Categories B, N, S.**

51) J. P. LeBlanc, K. Dogancay, R. A. Kennedy, and C. R. Johnson, Jr., "Effects of input data correlation on the convergence of blind adaptive equalizers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Adelaide, SA, Australia, Apr. 19–22, 1994, pp. 313–316. **Categories B, R, S.**

52) J. P. LeBlanc, I. Fijalkow, B. Huber, and C. R. Johnson, Jr., "Fractionally spaced CMA equalizers under periodic and correlated inputs," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, May 9–12, 1995, pp. 1041–1044. **Categories F, S, T.**

53) J. P. LeBlanc, "Effects of source distributions and correlation on fractionally spaced blind constant modulus algorithm equalizers," Ph.D. dissertation, Cornell University, Ithaca, NY, 1995. **Categories F, S, T, R.**

54) J. P. LeBlanc and S. W. McLaughlin, "Non-equiprobable constellation shaping and blind constant modulus algorithm equalization," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 1996, pp. 901–903. **Categories F, S.**

55) J. P. LeBlanc, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally spaced constant modulus algorithm blind equalizer error surface characterization: Effects of source distributions," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Atlanta, GA, May 7–9, 1996, pp. 2944–2947. **Categories F, S, T, R.**

56) J. P. LeBlanc and C. R. Johnson, Jr., "Global CMA error surface characteristics, source statistic effects: Polytopes and manifolds," in *Proc. Int. Conf. Digital Signal Processing*, Santorini, Greece, July 2–4, 1997, pp. 131–134. **Categories F, S, T, P.**

57) J. P. LeBlanc, I. Fijalkow, and C. Richard Johnson, Jr., "CMA fractionally spaced equalizers: Stationary points and stability under IID and temporally correlated sources," *Int. J. Adaptive Contr. and Signal Processing*, vol. 12, no. 2, pp. 135–155, March 1998. **Categories F, S, T.**

58) Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard equalizers," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 31–Nov. 2, 1994, vol. 1, pp. 617–621. **Categories F, T, P.**

59) ——, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 43, pp. 2120–2129, Sept. 1995. **Categories B, T, L, I.**

60) ——, "Global convergence of fractionally spaced Godard (CMA) adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 818–826, Apr. 1996. **Categories F, T, P.**

61) Y. Li, K. J. R. Liu, and Z. Ding, "Length and cost dependent local minima of unconstrained blind channel equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 2726–2735, Nov. 1996. **Categories B, T, L.**

62) Y. Li and K. J. R. Liu, "Static and dynamic convergence behavior of adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 22, pp. 2736–2745, Nov. 1996. **Categories B, L, E.**

63) F. Lopez de Victoria, A. Bosser, I. Fijalkow, C. R. Johnson, Jr., and J. R. Treichler, "Observed (mis)behavior of CMA with periodic sources: Assessment and guidelines," in *Proc. IEEE Signal Processing Workshop*, Yosemite National Park, CA, Oct. 2–5, 1994, pp. 261–264. **Categories B, S.**

64) W. E. Meyer and J. P. LeBlanc, "Blind adaptive fractionally spaced CMA in the presence of channel noise," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 1996, pp. 373–374. **Categories F, N.**

65) C. Papadias, "On the existence of undesired global minima of Godard equalizers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Munich, Germany, Apr. 20–24, 1997, pp. 3937–3940. **Categories F, T, S, P.**

66) J. J. Shynk and C. K. Chan, "A comparative analysis of the stationary points of the constant modulus algorithm based on Gaussian assumptions," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 1249–1252. **Categories B, S, T.**

67) ——, "Error surfaces of the constant modulus algorithm," in *Int. Symposium Circuits and Systems*, New Orleans, LA, May 1990, pp. 1335–1338. **Categories B, T, S.**

68) ——, "Performance surfaces of the constant modulus algorithm based on a conditional Gaussian model," *IEEE Trans. Signal Processing*, vol. 41, pp. 1965–1969, May 1993. **Categories B, S, T.**

69) J. O. Smith and B. Freidlander, "Global convergence of the constant modulus algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Tampa, FL, Mar. 26–29, 1985, pp. 1161–1164. **Categories B, P, T.**

70) R. Swaminathan and J. K. Tugnait, "On improving the convergence of constant modulus algorithm adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, Apr. 27–30, 1993, pp. 340–343. **Categories B, I.**

71) L. Tong and H. Zeng, "Channel surfing re-initialization for the constant modulus algorithm," *IEEE Signal Processing Lett.*, vol. 4, pp. 85–87, Mar. 1997. **Categories F, I.**

72) A. Touzni, I. Fijalkow, and J. R. Treichler, "Fractionally spaced CMA under channel noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Atlanta, GA, May 7–9, 1996, pp. 2674–2677. **Categories F, N, T.**

73) ——, "Robustness of fractionally spaced equalization by CMA to lack of channel disparity," in *Proc. IEEE Signal Processing Workshop Statistical Signal Array Processing*, Corfu, Greece, June 24–26, 1996, pp. 144–147. **Categories F, D, T.**

74) A. Touzni and I. Fijalkow, "Does fractionally spaced CMA converge faster than LMS?" in *Proc. European Signal Processing Conf.*, Trieste, Italy, Sept. 1996, pp. 1227–1230. **Categories F, R, T, P.**

75) ——, "Channel robust blind fractionally spaced equalization," in *IEEE Signal Signal Processing Workshop Signal Processing Advances Wireless Commun.*, Paris, France, Apr. 1997, pp. 33–36. **Categories F, D, I.**

76) ——, "Robustness of blind fractionally spaced identification/equalization to loss of channel disparity," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Munich, Germany, Apr. 1997, pp. 3937–3940. **Categories F, D, T.**

77) J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol ASSP-31, pp. 459–472, Apr. 1983. **Categories B, T, I, P.**

78) J. R. Treicher and M. G. Larimore, "Convergence rates for the constant modulus algorithm with sinusoidal inputs," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Tampa, FL, Mar. 26–29, 1985, pp. 1157–1160. **Categories B, R, S.**
79) ——, "The tone capture properties of CMA-based interference suppressors," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP–33, pp. 946–958, Aug. 1985. **Categories B, S.**
80) ——, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, pp. 420–431, Apr. 1985. **Categories B, F.**
81) J. R. Treichler, V. Wolff, and C. R. Johnson, Jr., "Observed misconvergence in the constant modulus adaptive algorithm," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1991, pp. 663–667. **Category S.**
82) J. R. Treichler, L. Tong, I. Fijalkow, C. R. Johnson, Jr., and C. U. Berg, "On the current shape of FSE-CMA behavior theory," in *Proc. IEEE Signal Processing Workshop Signal Processing Advances Wireless Communications*, Paris, France, Apr. 1997, pp. 105–108. **Categories F, N, L.**
83) J. K. Tugnait, "A parallel multimodel: CMA/Godard adaptive filter bank approach to fractionally spaced blind adaptive equalization," in *Proc. IEEE Int. Conf. Communications*, New Orleans, LA, May 1–5, 1994, vol. 1, pp. 549–553. **Category F.**
84) ——, "On fractionally spaced blind adaptive equalization under symbol timing offsets using Godard and related equalizers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, May 9–12, 1995, vol. 3, pp. 1976–1979. **Category F.**
85) ——, "On fractionally spaced blind adaptive equalization under symbol timing offsets using Godard and related equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 1817–1821, July 1996. **Categories B, F.**
86) S. Vembu, S. Verdu, R. A. Kennedy, and W. A. Sethares, "Convex cost functions in blind equalization," *IEEE Trans. Signal Processing*, vol. 42, pp. 1952–1960, Aug. 1994. **Category B.**
87) M. Wu and F. Cornett, "Discrete-time and continuous-time constant modulus algorithm analysis," in *Proc. Southeastern Symp. Systems Theory*, Starkville, MS, Mar. 12–14, 1995, pp. 504–508. **Categories B, R, E.**
88) H. H. Zeng and L. Tong, "On the performance of CMA in the presence of noise—Some new results on blind channel estimation: Performance and algorithms," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 1996, pp. 890–894. **Categories F, N, T.**
89) H. H. Zeng, L. Tong, and C. R. Johnson, Jr., "Behavior of fractionally spaced constant modulus algorithm: Mean square error, robustness and local minima," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1996, pp. 305–309. **Categories F, N, T, D, L.**
90) H. H. Zeng and L. Tong, "The MSE performance of constant modulus receivers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Munich, Germany, Apr. 1997, pp. 3577–3580. **Categories B, F, N, T.**
91) H. H. Zeng, L. Tong, and C. R. Johnson, Jr., "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1523–1538, July 1998. **Categories F, N, T, D, L.**
92) S. Zeng, H. H. Zeng, and L. Tong, "Blind equalization using CMA: Performance analysis and a new algorithm," in *Proc. IEEE Int. Conf. Communications*, Dallas, TX, June 1996, pp. 847–851. **Categories F, N, T.**
93) E. Zervas, J. G. Proakis, and V. Eyuboglu, "Effect of constellation shaping on blind equalization," in *Proc. Int. Soc. Optical Engineers*, San Diego, CA, pp. 1393–1397, 1991. Categories **B, S.**

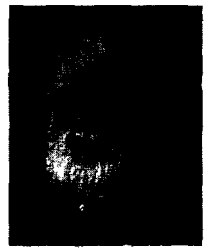
ACKNOWLEDGMENT

The authors would like to thank the following people for their comments on earlier versions of this paper: C. Papadias, J. Treichler, I. Fijalkow, A. Sayed, D. Jones, A. M. Baksho, W. Chung, J. Balakrishnan, M. A. Poubelle, C. P. Hankey, and A. F. Bob.

REFERENCES

[1] J. W. M. Bergmans and A. J. E. M. Janssen, "Robust data equalization, fractional tap spacing and the Zak transform," *Philips J. Res.*, vol. 42, no. 4, pp. 351–398, 1987.
[2] Z. Ding, "Characteristics of band-limited channels unidentifiable from second-order cyclostationary statistics," *IEEE Signal Processing Lett.*, vol. 3, pp. 150–152, May 1996.
[3] G. D. Forney, L. Brown, M. V. Eyuboglu, and J. L. Moran III, "The V.34 high-speed modem standard," *IEEE Commun. Mag.*, vol. 34, pp. 28–33, Dec. 1996.
[4] R. D. Gitlin, "Fractionally spaced equalization: An improved digital transversal equalizer," *Bell Syst. Tech. J.*, vol. 60, pp. 275–296, Feb. 1981.
[5] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communications Principles.* New York: Plenum Press, 1992.
[6] S. Haykin, *Adaptive Filter Theory,* 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[7] C. R. Johnson, Jr. *et al.*, "On fractionally spaced equalizer design for digital microwave radio channels," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 1995, pp. 290–294.
[8] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.
[9] M. G. Larimore and M. J. Goodman, "Implementation of the constant modulus algorithm at RF bandwidths," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1985, pp. 626–630.
[10] E. A. Lee and D. G. Messerschmitt, *Digital Communication,* 2nd ed. Boston, MA: Kluwer, 1994.
[11] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell Syst. Techn. J.*, vol. 45, pp. 255–286, Feb. 1966.
[12] ——, "Signal filtering with the transversal equalizer," in *Proc. Allerton Conf. Circuits Syst. Theory,* Allerton, IL, Oct. 1969, p. 792.
[13] D. G. Luenberger, *Optimization by Vector Space Methods,* 2nd ed. New York: Wiley, 1990.
[14] O. Macchi and L. Guidoux, "A new equalizer: The double-sampling equalizer," *Annales des Telecommun.*, vol. 30, nos. 9/10, pp. 331–338, Sept. 1975.
[15] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing,* vol. 43, pp. 516–525, Feb. 1995.
[16] J. G. Proakis, *Digital Communications,* 3rd ed. New York: McGraw-Hill, 1995.
[17] S. U. H. Qureshi, "Adjustment of the position of the reference tap of an adaptive equalizer," *IEEE Trans. Commun.*, pp. 1046–1052, Sept. 1973.
[18] ——, "Adaptive equalization," *Proc. IEEE,* vol. 73, pp. 1349–1387, Sept. 1985.
[19] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulations systems," *IEEE Trans. Commun.*, pp. 679–682, June 1975.
[20] G. Strang, *Linear Algebra and its Applications,* 3rd ed. Fort Worth, TX: Harcourt Brace Jovanovich, 1988.
[21] L. Tong, "A fractionally spaced adaptive blind equalizer," in *Proc. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1992, pp. 711–716.
[22] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inform. Theory,* vol. 41, pp. 329–334, Jan. 1995.
[23] J. R. Treichler, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally spaced equalizers: How long should they really be?," *IEEE Signal Processing Mag.*, vol. 13, pp. 65–81, May 1996.
[24] J. R. Treichler, M. G. Larimore, and J. C. Harp, "Practical blind demodulators for high-order QAM signals," this issue, pp. 1907–1926.
[25] J. K. Tugnait, "On blind identifiability of multipath channels using fractional sampling and 2nd-order cyclostationary statistics," *IEEE Trans. Inform. Theory,* vol. 41, pp. 308–311, Jan. 1995.
[26] G. Ungerboeck, "Fractional tap spacing equalizer and consequences for clock recovery in data modems," *IEEE Trans. Commun.*, vol. 24, no. 8, pp. 856–864, Aug. 1976.

[27] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Engelwood Cliffs, NJ: Prentice-Hall, 1985.
[28] V. G. Wolff, R. P. Gooch, and J. R. Treichler, "Specification and development of an equalizer-demodulator for wideband digital microwave radio signals," in *Proc. IEEE Military Commun. Conf.*, San Diego, CA, Oct. 1988, pp. 461–467.
[29] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980.
[30] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 31, pp. 459–72, Apr. 1983.
[31] G. J. Foschini, "Equalizing without altering or detecting data (digital radio systems)," *AT&T Tech. J.*, vol. 64, no. 8, pp. 1885–1911, Oct. 1985.
[32] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard (CMA) adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 44, no. 4, pp. 818–826, Apr. 1996.
[33] I. Fijalkow, A. Touzni, and J. R. Treichler, "Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 56–66, Jan. 1997.
[34] M. G. Larimore and J. R. Treichler, "Convergence behavior of the constant modulus algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Boston, MA, Apr. 14–16, 1983, pp. 13–16.
[35] S. Lambotharan, J. Chambers, and C. R. Johnson, Jr., "Attraction of saddles and slow convergence in CMA adaptation," *Signal Processing*, vol. 59, no. 2, pp. 335–340, June 1997.
[36] A. Touzni and I. Fijalkow, "Does fractionally spaced CMA converge faster than LMS?," in *Proc. European Signal Processing Conf.*, Trieste, Italy, Sept. 1996, pp. 1227–1230.
[37] I. Fijalkow, C. Manlove, and C. R. Johnson, Jr., "Adaptive fractionally spaced blind CMA equalization: Excess MSE," *IEEE Trans. Signal Processing*, vol. 46, pp. 227–231, Jan. 1998.
[38] C. R. Johnson, Jr. and B. D. O. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary case," *Int. J. Adaptive Contr. Signal Processing*, vol. 9, pp. 301–324, July/Aug. 1995.
[39] J. P. LeBlanc, I. Fijalkow, and C. Richard Johnson, Jr., "CMA fractionally spaced equalizers: Stationary points and stability under IID and temporally correlated sources," *International Journal of Adaptive Contr. Signal Processing*, 1998, to appear.
[40] H. H. Zeng, L. Tong, and C. R. Johnson, Jr., "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1523–1538, July 1998.
[41] W. Chung and J. P. LeBlanc, "The local minima of fractionally spaced CMA blind equalizer cost function in the presence of channel noise," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Seattle, WA, May 1998, pp. 3345–3348.
[42] Y. Li and K. J. R. Liu, "Static and dynamic convergence behavior of adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 2736–2745, Nov. 1996.
[43] T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr., and M. Green, "On the robustness of the fractionally spaced constant modulus criterion to channel order undermodeling: Part I," in *Proc. IEEE Signal Processing Workshop on Signal Processing Advances Wireless Commun.*, Paris, France, Apr. 16–18, 1997, pp. 37–40.
[44] ——, "On the robustness of the fractionally spaced constant modulus criterion to channel order undermodeling: Part II," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Proc.*, Munich, Germany, Apr. 20–24, 1997, pp. 3605–3608.
[45] J. R. Treichler, V. Wolff, and C. R. Johnson, Jr., "Observed misconvergence in the constant modulus adaptive algorithm," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1991, pp. 663–667.
[46] R. A. Axford, Jr., L. B. Milstein, and J. R. Zeidler, "The effects on PN sequences on the misconvergence of the constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 46, pp. 519–523, Feb. 1998.
[47] J. P. LeBlanc, "Effects of source distributions and correlation on fractionally spaced blind constant modulus algorithm equalizers," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1995.



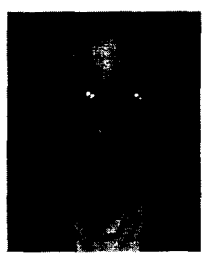
C. Richard Johnson, Jr. (Fellow, IEEE) was born in Macon, GA, in 1950. He received the Ph.D. degree in electrical engineering with minors in engineering-economic systems and art history from Stanford University, Stanford, CA, in 1977.

He is currently a Professor of Electrical Engineering and a member of the Graduate Field of Applied Mathematics at Cornell University, Ithaca, NY. His current research interest is in adaptive parameter estimation theory, which is useful in applications of digital signal processing to telecommunication systems. His principal focus in the 1990's has been blind linear equalization for intersymbol interference removal from received QAM sources.

Philip Schniter was born in Evanston, IL, in 1970. He received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993. Since 1996 he has been working toward the Ph.D. degree in electrical engineering at Cornell University, Ithaca, NY

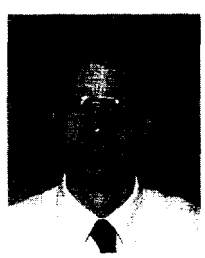
From 1993 to 1996 he was employed by Tektronix, Inc., Beaverton, OR, as a Systems Engineer. While there he worked on signal processing aspects of video and communications instrumentation design, including algorithms, software, and hardware architectures. His research interests are in signal processing, communications, and control, and include blind adaptive equalization.

Mr. Schniter is a recipient of the Schlumberger Fellowship.



Thomas J. Endres (Member, IEEE) received the B.S. degree from Cornell University, Ithaca, NY, in 1990, the M.S. degree from the University of Southern California, Los Angeles, 1994, and the Ph.D. degree from Cornell University in 1997, all in electrical engineering.

From 1990 to 1994, he was employed by Hughes Space and Communications, Los Angeles, CA. Since 1997 he has been a Founding Member of Sarnoff Digital Communications, Newtown, PA. His current research interests include blind equalization and adaptive systems.

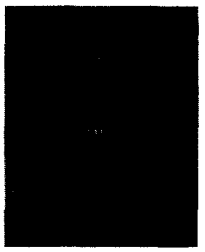


James D. Behm received the M.S. degree in mathematics in 1975 from the University of Arizona, Tucson, and the M.S. degree in computer science in 1986 from Johns Hopkins Whiting School of Engineering, Laurel, MD.

Since 1975, he has been a Mathematician with the Department of Defense, Ft. Meade, MD. In Spring, 1997 he was a Visiting Researcher at Cornell University, Ithaca, NY. His current research interests include signal processing and communications.

Donald R. Brown received the B.S. and M.S. degrees in electrical engineering from the University of Connecticut, Storrs, in 1992 and 1996, respectively. He is currently a graduate student at Cornell University, Ithaca, NY.

From 1992 to 1997 he worked for the General Electric Company, Plainville, CT, as a Development Engineer. His research interests include adaptive signal processing and communications.



Raúl A. Casas was born in Denver, CO, in 1972. He received the B.S. and M.S. degrees in electrical engineering from Cornell University, Ithaca, NY, in 1994 and 1996, respectively. He is currently pursuing the Ph.D. degree at Cornell University.

His thesis work is a result of collaborative research with the United Technologies Research Center on identification and control of nonlinear combustion chamber dynamics. His current research interests include the analysis of signal processing, communications, and control related applications from a nonlinear dynamical systems perspective.

Mr. Casas is a recipient of the National Science Foundation Minorities Fellowship.